



# THÈSE

En vue de l'obtention du

**DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE**

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

---

---

Présentée et soutenue le 5 Juillet 2018 par :

**VALENTIN BAHIER**

**Spectre de matrices de permutation aléatoires**

---

---

## JURY

FLORENT BENAYCH-GEORGES  
MIREILLE CAPITAINE  
DJALIL CHAFAÏ  
REDA CHHAIBI  
CATHERINE DONATI-MARTIN  
LAURENT MICLO  
JOSEPH NAJNUDEL  
ALAIN ROUAULT

Université Paris Descartes  
Université Toulouse 3  
Université Paris-Dauphine  
Université Toulouse 3  
Université de Versailles  
Université Toulouse 3  
Université de Cincinnati  
Université de Versailles

Rapporteur  
Examineur  
Examineur  
Examineur  
Rapporteur  
Examineur  
Directeur de thèse  
Examineur

---

**École doctorale et spécialité :**

*MITT : Domaine Mathématiques : Mathématiques appliquées*

**Unité de Recherche :**

*Institut de Mathématiques de Toulouse (UMR 5219)*

**Directeur de Thèse :**

*Joseph Najnudel*

**Rapporteurs :**

*Florent Benaych-Georges et Catherine Donati-Martin*



*Vers l'infini, et au-delà...*



# Remerciements

Tout d'abord, je tiens à remercier énormément mon directeur de thèse Joseph Najnudel. Depuis mon stage de Master 2 sous sa direction, il s'est toujours rendu très disponible pour répondre à mes questions et me guider, même avec la distance géographique entre nous qui ne facilitait pas nos communications ces deux dernières années. Je garde d'excellents souvenirs de mes deux séjours aux USA pour travailler auprès de lui. Merci beaucoup pour l'accueil et tous les bons moments passés ensemble, avec de riches rencontres là-bas : merci Nancy, Wlodek, Scott, Yizao, Nag, Costel et Magda, Robert, Fabrice, et tout particulièrement Ju-Yi de m'avoir si bien intégré.

Je remercie vivement Mireille Capitaine, en partie grâce à qui j'ai pu commencer une thèse. Du début jusqu'à la fin elle a veillé à la cohérence de mon projet de thèse, et m'a laissé grande ouverte la porte de son bureau en cas de questions ou problèmes.

Je suis très reconnaissant envers Florent Benaych-Georges et Catherine Donati-Martin d'avoir accepté d'écrire un rapport sur mon travail, et pour leur contribution à l'amélioration de ce manuscrit. Un grand merci aussi aux autres membres du jury Djilil Chafaï, Reda Chhaïbi, Laurent Miclo et Alain Rouault de me faire l'honneur de leur présence.

J'ai aussi une pensée au personnel de l'IMT sans qui mes conditions de travail et mes joies administratives n'auraient pas été si bonnes, particulièrement je remercie Françoise, Martine, Marie-Line, Agnès, Delphine et Tamara.

Merci à tous les doctorants de l'IMT, en commençant par la fine équipe de doctorants ESP : Antoine, président des doctorants et meilleur grimpeur du tour d'Oléron vélo; Pierre, spécialiste en pâte à crêpes et des danses en tutu; Hugo, marcheur quantique ouvert et fourgonneur express des retours de soirées; William, le Booba des maths partisan de la convexité; Fabien, roi de la claquette et des mauvais calembours; je souligne aussi l'arrivage d'un excellent cru ESP très prometteur : Eva, Eyal, Gabriel, José, Camille, Valentin, Thien et Phong. Merci à vous !

Merci aussi à mes camarades du bureau 205, bureau que les autres ont peut-être trop souvent confondu avec la salle de pause : merci à Tatiana et Stéphane pour leur accueil et nombreux conseils, An pour sa bonne humeur et ses partages de mets vietnamiens savoureux, et Kiki pour son goût pour le gras et les pintes de vin blanc.

Je remercie ensuite mes adversaires aux championnats du monde de coinche de l'IMT Anton, Aurore, Kévin, Guillaume et Ioana, et aussi pour d'autres jeux de société et chouettes soirées Maxime, Laura et Damien.

Merci enfin aux autres doctorants avec qui j'ai partagé des bons moments : Baptiste, Jibar, Silvère, Jade, Jules, Julie, Maria, JP, Sourav, Ibrahim, Kamilia, Jean-Marc, Joachim, Florian, Alexis, Susely, Leyth, Anthony; et les anciens ESP Laure, Clément, Claire, Benoît, Kévin, Raph, Yuriy, Fanny, Sofiane, Sylvain, Malika; ainsi que les post-docs Magali et Tristan.

C'est aussi l'occasion pour moi de remercier les nombreux enseignants qui m'ont donné goût aux mathématiques et à l'enseignement depuis le collège jusqu'ici : mesdames et messieurs Perlemoine, Masmoudi, Gilardi, Ilan, Brault, Tougeron, Dutrieux, Delale, Pierre, Gradinaru, Breton, Cadre, et Genieys.

Depuis tout petit la musique a constitué pour moi un moteur de bonne humeur et de riches rencontres, dont beaucoup d'amitiés solides parmi lesquelles comptent Paul, Mathilde, Romain, Alexis, Marich, Manu; et ceux que je commence à voir moins souvent mais que je n'oublie pas Adrien, Benjamin, Thibault, Sophie, Mathilde, Caroline... Je vous embrasse !

Merci aussi à la fanfare des Frères Pouetards pour tous ces génialissimes plans musicaux ensemble depuis trois ans. Quel régal d'être parmi vous !

Je profite de ces quelques lignes pour émettre des pensées vers trois amis qui me sont chers et qui ont marqué mon parcours scolaire : mon lapin gratteux biéristo-boucanier Thomas, mon pote poète métalleux Rodg', et mon maître arithméticien poker-buveur Phi. Bises les gars !

Enfin, je n'aurais jamais pu aller aussi loin dans des études sans eux, je veux bien sûr parler des membres de ma famille.

Merci à mes parents pour leur soutien tout au long de ma vie étudiante et leur fierté de me voir poursuivre dans cette voie. J'ai beaucoup de chance qu'ils soient aussi exemplaires dans tout ce qu'il font, et particulièrement dans leur travail. Merci de m'avoir transmis les valeurs du courage et du travail bien fait, du respect, de l'esprit d'équipe et du plaisir de jouer ensemble.

Merci à ma grande sœur d'être fière de ce que je fais, je suis touché par l'intérêt qu'elle porte envers mes activités mathématiques ou autres. C'est un bonheur d'avoir grandi auprès d'elle. Merci de me faire rire en me considérant comme l'autiste de la famille.

Merci à mes cousins, cousines, oncles, tantes et grands-parents, pour tous ces bons moments passés ensemble chaque fois que l'on se voit.

Merci à la famille de Maylis que je suis très heureux de connaître depuis quelques années déjà. C'est un réel plaisir de venir toutes ces fois à CLF pour discuter, jouer, manger des biscuits, se reposer, et s'aérer l'esprit avec des travaux manuels riches et variés. Je touche du foin pour que cela continue.

Merci enfin à ma petite muce qui me comble d'amour et de joies au quotidien. <3

# Table des matières

<b>1</b>	<b>Introduction générale</b>	<b>1</b>
1.1	Les permutations aléatoires . . . . .	1
1.1.1	Loi uniforme . . . . .	2
1.1.2	Lois d'Ewens . . . . .	3
1.1.3	Lois plus générales . . . . .	5
1.2	Matrices de permutation . . . . .	6
1.2.1	Lien entre valeurs propres et structure en cycles . . . . .	6
1.2.2	Répartition des valeurs propres . . . . .	8
1.2.3	Statistiques linéaires . . . . .	8
1.2.4	Polynômes caractéristiques et fonctions centrales . . . . .	10
1.3	Matrices de permutation modifiées . . . . .	13
1.3.1	Similarités avec le CUE . . . . .	14
1.3.2	Répartition des valeurs propres . . . . .	15
1.3.3	Statistiques linéaires . . . . .	16
1.3.4	Polynômes caractéristiques et fonctions centrales . . . . .	18
1.4	Présentation des résultats . . . . .	19
1.4.1	Échelles macroscopique et mésoscopique . . . . .	20
1.4.2	Échelle microscopique . . . . .	23
<b>2</b>	<b>Counting function of eigenvalues at macroscopic and mesoscopic scales</b>	<b>31</b>
2.1	Introduction . . . . .	31
2.1.1	Random permutation matrices . . . . .	31
2.1.2	Notations and main results . . . . .	33
2.2	Preliminaries . . . . .	35
2.2.1	Cesàro means of fractional order . . . . .	35
2.2.2	Feller Coupling . . . . .	38
2.3	Proof of Theorem 1 . . . . .	40
2.3.1	Mean and variance . . . . .	40
2.3.2	Limiting normality for a finite number of fixed arcs . . . . .	44
2.4	Proof of Theorem 2 . . . . .	51
2.4.1	Variance . . . . .	51
2.4.2	Limiting normality for a single mesoscopic arc . . . . .	59
2.5	Spacing between eigenvalues . . . . .	62
2.5.1	Largest spacing between two consecutive distinct eigenvalues . . . . .	62
2.5.2	Smallest spacing between two consecutive distinct eigenvalues . . . . .	63

<b>3</b>	<b>Counting function of eigenvalues at microscopic scale</b>	<b>77</b>
3.1	Introduction . . . . .	77
3.1.1	Spectrum of random permutation matrices . . . . .	77
3.1.2	Main results and outline of the chapter . . . . .	82
3.2	Two natural limiting counting functions . . . . .	84
3.3	Limiting point process related to permutation matrices with modification . . . . .	87
3.4	Continuous analog of the Feller coupling . . . . .	93
3.5	Limiting point process related to permutation matrices . . . . .	98
3.5.1	Approximation of $S$ by $T$ . . . . .	99
3.5.2	Proof of Theorem 3.1.7 (i) . . . . .	100
3.5.3	Proof of Theorem 3.1.7 (ii) . . . . .	103
3.6	Translation of the limiting point process related to permutation matrices . . . . .	106
<b>4</b>	<b>Characteristic polynomials</b>	<b>115</b>
4.1	Introduction . . . . .	115
4.1.1	Convergence of characteristic polynomials . . . . .	115
4.1.2	Generating random permutations . . . . .	116
4.1.3	Main results and outline of the chapter . . . . .	122
4.2	Proof of the main theorem . . . . .	124
4.2.1	Quotient of characteristic polynomials related to modified permutation matrices . . . . .	124
4.2.2	Quotient of characteristic polynomials related to permutation matrices . . . . .	128
4.3	Properties of the limiting function . . . . .	134
4.4	More general central measures . . . . .	137
4.4.1	Proof of Theorem 4.4.2 (i) . . . . .	139
4.4.2	Proof of Theorem 4.4.2 (ii) . . . . .	142
4.4.3	Properties of the limiting functions . . . . .	143
<b>5</b>	<b>Conclusion et perspectives</b>	<b>145</b>



# Notations

La liste suivante présente les principales notations utilisées dans ce manuscrit.

## Paramètres et ensembles

$\theta$	Paramètre de la loi d'Ewens.
$I, I_k$	Intervalle du cercle unité.
$J$	Intervalle de l'ensemble des réels.
$S^1$	Cercle unité.
$\mathfrak{S}_n$	$n$ -ème groupe symétrique.
$S^1 \wr \mathfrak{S}_n$	Produit en couronne ( <i>wreath product</i> en anglais) des groupes $S^1$ et $\mathfrak{S}_n$ .
$\nabla$	$\left\{ (x_1, x_2, \dots) \in [0, 1]^\infty : x_1 \geq x_2 \geq \dots, \sum_{j=1}^{+\infty} x_j \leq 1 \right\}$ .
$\nabla'$	$\left\{ (x_1, x_2, \dots) \in [0, 1]^\infty : x_1 \geq x_2 \geq \dots, \sum_{j=1}^{+\infty} x_j = 1 \right\}$ .

## Variables aléatoires

$M_n$	Matrice de permutation de taille $n$ .
$\widetilde{M}_n$	Matrice de permutation modifiée de taille $n$ .
$X_n^I, \widetilde{X}_n^I$	Nombre de valeurs propres de $M_n$ (resp. $\widetilde{M}_n$ ) appartenant à $I$ .
$a_{n,j}$	Nombre de cycles de taille $j$ dans une permutation aléatoire de $[n]$ .
$W_j$	Variable de Poisson de paramètre $\theta/j$ .
$\Phi_j, \Phi_{j,k}$	Variables uniformes sur $[0, 1]$ .
$\mathcal{D}_n, \widetilde{\mathcal{D}}_n$	Plus grande distance entre deux valeurs propres consécutives d'un élément aléatoire de $\mathfrak{S}_n$ (resp. $S^1 \wr \mathfrak{S}_n$ ).
$d_n, \widetilde{d}_n$	Plus petite distance entre deux valeurs propres distinctes consécutives d'un élément aléatoire de $\mathfrak{S}_n$ (resp. $S^1 \wr \mathfrak{S}_n$ ).
$L_{n,1}$	Plus grande taille de cycle d'un élément aléatoire de $\mathfrak{S}_n$ .
$(A_{n,1}, \dots, A_{n,n})$	Vecteur des tailles de cycles dans l'ordre d'apparition (ordre croissant des cycles selon le plus petit élément qu'ils contiennent).
$\ell_{n,j}$	Longueur du $j$ -ème cycle d'une permutation de $[n]$ (chapitre 3). Variable binomiale de paramètres $n$ et $y_j$ (chapitre 4).
$y_j^{(n)}$	Longueur normalisée $y_j^{(n)} = \frac{\ell_{n,j}}{n}$ .
$\tau_n, \widetilde{\tau}_n$	Mesure empirique des angles propres de $M_n$ (resp. $\widetilde{M}_n$ ), normalisés de telle façon à avoir des espacements moyens de 1.
$\tau_\infty, \widetilde{\tau}_\infty$	Versions $n = \infty$ de $\tau_n$ et $\widetilde{\tau}_n$ .
$\mathcal{X}$	Processus de Poisson d'intensité $\frac{\theta}{x} dx$ sur $]0, +\infty[$ .

## Lois et convergences

$\mathbb{P}_\theta^{(n)}$	Loi d'Ewens de paramètre $\theta$ sur $\mathfrak{S}_n$ . $\mathbb{P}_\theta^{(n)}(\sigma) = \frac{\theta^{K(\sigma)}}{\theta(\theta+1)\cdots(\theta+n-1)}$ .
$\xrightarrow{a.s.}$	Convergence presque sûre.
$\xrightarrow{\mathbb{P}}$	Convergence en probabilité.
$\xrightarrow{d}$	Convergence en loi.
$\implies$	Convergence en loi sur l'espace des fonctions continues de $\mathbb{C}$ dans $\mathbb{C}$ muni de la topologie de la convergence uniforme sur tout compact.
$\mathbb{P}_X(A)$	Espérance conditionnelle de $\mathbb{1}_A$ (indicatrice de l'événement $A$ ) sachant la variable aléatoire $X$ .

### Quantités liées aux moyennes de Cesàro

$A_n^\delta$	Nombre de Cesàro d'ordre $\delta$ . $A_n^\delta = \frac{\Gamma(n+\delta+1)}{\Gamma(1+\delta)n!}$ .
$\sigma_n^\theta(w)$	Moyenne de Cesàro d'ordre $\theta$ de la suite $w = (w_j)_{j \geq 0}$ . $\sigma_n^\theta(w) = \sum_{j=0}^n \frac{A_{n-j}^{\theta-1}}{A_n^\theta} w_j$ .
$\Psi_n(j)$	$\Psi_n(j) = \frac{\Gamma(n+1)\Gamma(\theta+n-j)}{\Gamma(n-j+1)\Gamma(\theta+n)}$ .
$(C, \theta)$	Convergence au sens de Cesàro d'ordre $\theta$ .

### Autres symboles

$\lfloor x \rfloor, \lceil x \rceil, \{x\}$	Partie entière de $x$ , partie entière supérieure de $x$ , partie fractionnaire de $x$ . $x = \lfloor x \rfloor + \{x\} = \lceil x \rceil - \{-x\}$ .
$\ x\ $	Distance de $x$ au plus proche entier. $\ x\  = \min(\{x\}, 1 - \{x\})$ .
$a = \mathcal{O}_x(b)$	
ou $a \ll_x b$	Il existe une constante $C_x$ dépendante de $x$ telle que $ a/b  \leq C_x$ .
$a \sim b$	$a/b \rightarrow 1$ .
$a = o(b)$	$a/b \rightarrow 0$ .
$[n]$	Ensemble des entiers de 1 à $n$ .
$\llbracket m, n \rrbracket$	Ensemble des entiers de $m$ à $n$ .
$\Re(z), \Im(z)$	Partie réelle de $z$ , partie imaginaire de $z$ .
$<.,.>$	Crochet de distribution.

# 1. Introduction générale

Dans les années 20, Wishart [90] introduit des matrices aléatoires pour étudier le comportement de plusieurs variables dans le cadre d’analyses multivariées en statistique. Il fournit des formules permettant d’estimer des coefficients de corrélation et de faire des tests statistiques sur des matrices de covariance empiriques, avec une loi qui porte aujourd’hui son nom. Les matrices aléatoires connaissent un second élan dans les années 50 en physique statistique et mécanique quantique avec les travaux de Wigner. Dans l’intention de simplifier les modèles d’étude de spectres nucléaires, Wigner suggère de remplacer certains opérateurs de dimension infinie dans les équations par des opérateurs de dimension finie respectant les mêmes symétries. Plus précisément, il prend comme hamiltoniens de grandes matrices aléatoires hermitiennes, puis cherche à déterminer les propriétés asymptotiques de ces matrices lorsque leur taille tend vers l’infini. Un des grands succès de son étude est le phénomène d’*universalité* qui se produit pour la loi limite non-triviale de la mesure spectrale empirique : pour de vastes ensembles de matrices aléatoires, celle-ci ne dépend pas de la loi des entrées de la matrice. Les matrices aléatoires ont dès lors commencé à susciter de l’intérêt dans de nombreux domaines de recherche tels que les systèmes intégrables, les télécommunications, la combinatoire, la théorie des graphes, les probabilités libres, les processus déterminantaux, et la théorie des nombres.

Dans ce dernier domaine par exemple, Montgomery [67] conjecture en 1972 que les zéros de la fonction zeta de Riemann sur l’axe critique  $\frac{1}{2} + it$  se comportent comme des angles propres de matrices unitaires suivant la mesure de Haar, le Circular Unitary Ensemble (CUE), c’est-à-dire de matrices unitaires de loi invariante par multiplication à droite ou à gauche par une matrice unitaire quelconque. En 2000, Keating et Snaith [57] font une avancée majeure en montrant que le polynôme caractéristique de matrices unitaires aléatoires évalué en des points du cercle unité serait un bon modèle pour décrire la fonction zeta sur l’axe critique.

Dans cette thèse nous nous intéressons à des ensembles de matrices unitaires aléatoires liées à des permutations aléatoires.

## 1.1 Les permutations aléatoires

Les permutations apparaissent naturellement dans de nombreux phénomènes, et nous permettent d’exprimer des façons de réordonner des objets entre eux. Leurs structures combinatoires ont été beaucoup étudiées et comprises ces cinquante dernières années, notamment grâce à l’introduction de concepts probabilistes.

Par structures combinatoires des permutations on veut souvent faire référence à leur structure en cycles, définie de la manière suivante :

Toute permutation  $\sigma$  de  $[n]$  peut être décomposée en cycles à supports disjoints, et cette écriture est unique à permutation des cycles près. Les nombres de cycles de mêmes tailles dans cette décomposition constituent un  $n$ -uplet  $(a_1, \dots, a_n)$ , appelé **structure en cycles** de  $\sigma$ , où  $a_j$  est le nombre de cycles de taille  $j$  dans la décomposition de  $\sigma$  en cycles à supports disjoints,  $1 \leq j \leq n$ .

La théorie des probabilités donne un cadre de travail simple et rigoureux pour l'étude de certaines propriétés des permutations en particulier lorsque  $n$  tend vers l'infini. Cela se traduit tout d'abord par la considération d'une loi de tirage sur l'ensemble des permutations de  $[n]$ , c'est-à-dire d'une mesure de probabilité sur le groupe symétrique  $\mathfrak{S}_n$ .

### 1.1.1 Loi uniforme

La loi de tirage sur  $\mathfrak{S}_n$  qui attribue un même poids à chacun de ses éléments est la loi uniforme : chaque permutation de  $[n]$  a autant de chance d'être tirée. Cette loi est en quelque sorte la plus naturelle pour aborder l'étude de  $\mathfrak{S}_n$  d'un point de vue global, au sens où la fréquence d'apparition d'une propriété donnée après  $k$  tirages tend vers la proportion d'éléments de  $\mathfrak{S}_n$  ayant cette propriété, lorsque  $k$  tend vers l'infini (par la loi des grands nombres).

Nous rappelons brièvement ci-après quelques résultats et grandes étapes relatives à l'étude des permutations sous la mesure uniforme.

En 1942-44, Goncharov [47] [46] étudie les nombres de cycles d'une taille donnée, le nombre total de cycles ainsi que la taille du plus grand cycle. Il établit notamment un résultat de normalité asymptotique pour le nombre total de cycles  $K_n$  d'une permutation aléatoire de  $[n]$  :

$$\frac{K_n - \ln n}{\sqrt{\ln n}} \longrightarrow \mathcal{N}(0, 1)$$

en loi lorsque  $n$  tend vers l'infini.

En 1959-64, Golomb [44] [45] montre que l'espérance de la plus grande longueur de cycle divisée par  $n$  admet une limite quand  $n$  tend vers l'infini et en donne une valeur approchée ( $\simeq 0.624329\dots$ ). Cette limite est connue sous le nom de *constante de Golomb*, ou encore *constante de Golomb-Dickman*, et apparaît aussi en théorie des nombres dans la décomposition d'entiers en facteurs premiers.

En 1966, Shepp et Llyod [81] poursuivent l'étude en s'intéressant aux tailles ordonnées des cycles. En particulier ils calculent formellement les moments asymptotiques des plus grands cycles et des plus petits cycles lorsque  $n$  tend vers l'infini.

En 1967, Erdős et Turán [38] établissent un résultat de normalité asymptotique pour l'ordre d'une permutation (*i.e.* le plus petit entier non nul  $d$  tel que  $\sigma^d = \text{id}$ ) : si  $O(\sigma_n)$  désigne l'ordre d'un élément aléatoire  $\sigma_n$  de  $\mathfrak{S}_n$ ,

$$\frac{O(\sigma_n) - \frac{1}{2}(\ln n)^2}{\sqrt{\frac{1}{3}(\ln n)^3}} \longrightarrow \mathcal{N}(0, 1)$$

en loi lorsque  $n$  tend vers l'infini.

En 1974, les travaux de Watterson [86] permettent en particulier de voir que les premiers moments factoriels joints des nombres de cycles de mêmes tailles coïncident avec les moments factoriels de variables de Poisson indépendantes de paramètres  $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$ .

Les résultats que nous venons de mentionner et beaucoup d'autres se trouvent dans l'excellent livre d'Arratia, Barbour et Tavaré [4], paru en 2003. Dans ce livre, d'autres structures combinatoires logarithmiques sont abordées et comparées, avec par exemple les décompositions d'entiers aléatoires en produit de facteurs premiers, ou les applications aléatoires de  $[n]$  dans lui-même. Pour en revenir aux permutations, Arratia, Barbour et Tavaré recueillent et développent un grand nombre de résultats pour des lois plus générales que la loi uniforme : la famille des lois d'Ewens. Nous motivons l'étude de ces lois dans la section suivante.

Aussi, depuis les années 1960, d'importants travaux ont été réalisés concernant les plus longues sous-suites croissantes (ou décroissantes) dans des permutations aléatoires. Leur étude peut permettre d'améliorer des algorithmes de tri. En effet, si dans une liste donnée on arrive à identifier une grande sous-suite croissante, il suffit de déplacer les autres éléments de cette liste et cela peut s'avérer plus efficace que par d'autres méthodes. Avec la correspondance de Robinson et Schensted [80], la plus longue sous-suite croissante correspond au nombre de colonnes (ou de lignes pour les sous-suites décroissantes) d'un des deux *tableaux de Young* de mêmes formes associés.

En 1999, Baik, Deift et Johansson [8] établissent que la taille  $l_n$  de la plus grande sous-suite croissante d'une permutation aléatoire de  $[n]$  vérifie

$$\frac{l_n - 2\sqrt{n}}{n^{1/6}} \longrightarrow Z$$

en loi lorsque  $n$  tend vers l'infini, où  $Z$  est une variable suivant la loi de Tracy-Widom [83] de paramètre  $\beta = 2$ . Il est remarquable d'obtenir la même loi limite que celle de la plus grande valeur propre renormalisée de matrices de l'ensemble gaussien unitaire (GUE).

Les permutations aléatoires suivant la loi uniforme apparaissent dans divers contextes. Par exemple, Albert, dans sa thèse [1], les utilise en statistiques pour faire des tests d'indépendance, dont l'approche par permutation a été initiée par Fisher [42] en 1935.

### 1.1.2 Lois d'Ewens

Les lois d'Ewens sont des déformations à un paramètre de la loi uniforme, paramètre noté usuellement  $\theta > 0$ , et tirent leur nom du mathématicien biologiste Warren Ewens. Depuis les années 60-70, Ewens analyse et développe de nombreux modèles en *génétique des populations*, dont celui-ci : considérons une population vérifiant certaines hypothèses (essentiellement on suppose la population de taille infinie en équilibre sans migration, sans sélection, et que les mutations donnent toujours de nouveaux allèles). Tirons aléatoirement  $n$  gamètes de cette population. Alors pour un certain locus fixé, la probabilité qu'il y ait  $a_1$  allèles en un exemplaire,  $a_2$  allèles en deux exemplaires, ..., et  $a_n$  allèles en  $n$  exemplaires est décrite par la loi d'Ewens de paramètre  $\theta$  sur les partitions de  $n$ , où  $\theta$

est le facteur de mutation, et s'écrit

$$p_\theta((a_1, \dots, a_n)) = \frac{n!}{\theta(\theta+1) \dots (\theta+n-1)} \prod_{j=1}^n \frac{\theta^{a_j}}{j^{a_j} a_j!}.$$

Cette formule célèbre est connue sous le nom d'*Ewens' sampling formula* et suscite beaucoup d'intérêt (voir par exemple [75], [49], [50], [3], [6] et [26]). Plus précisément, le paramètre  $\theta$  dans ce modèle appartient à  $]0, +\infty[$  et décrit la propension à donner lieu à de nouveaux allèles. Le cas limite  $\theta = 0$  correspond à l'absence totale de mutation, c'est-à-dire tous les individus de la population ont le même génotype et donc  $a_1 = a_2 = \dots = a_{n-1} = 0$  et  $a_n = 1$ , tandis que le cas limite  $\theta = +\infty$  correspond au fait que les allèles des individus sont deux à deux distincts et donc  $a_1 = n$  et  $a_2 = \dots = a_n = 0$ . De cette loi d'Ewens sur les partitions de  $n$  on déduit facilement la loi correspondante sur les permutations de  $[n]$ , encore appelée loi d'Ewens. En effet, par une formule de Cauchy, le nombre de permutations  $\sigma$  de  $[n]$  ayant pour structure en cycles  $(a_1, \dots, a_n)$  est

$$\frac{n!}{\prod_{j=1}^n j^{a_j} a_j!}$$

sous réserve que  $\sum_{j=1}^n j a_j = n$ , et donc la loi d'Ewens de paramètre  $\theta$  sur les permutations de  $[n]$  est décrite par les quantités

$$\frac{\theta^{K(\sigma)}}{\theta(\theta+1) \dots (\theta+n-1)}$$

où  $K(\sigma) = a_1 + \dots + a_n$  est le nombre total de cycles dans la décomposition en cycles à supports disjoints de  $\sigma$ . En particulier lorsque  $\theta = 1$  ces quantités valent  $\frac{1}{n!}$  et on retrouve donc la loi uniforme. Heuristiquement, plus  $\theta$  est grand, plus cela favorise l'apparition d'un grand nombre de cycles, et donc plus les cycles vont avoir tendance à être petits.

Un des outils appropriés dans ce cadre est la méthode de Feller qui consiste à donner une représentation de cette loi sur les permutations à partir de variables de Bernoulli  $\xi_j$  indépendantes de paramètres  $\frac{\theta}{\theta+j-1}$ . Dans le procédé de construction de ces permutations aléatoires, chaque événement  $\xi_j = 1$  correspond à fermer un cycle lorsqu'il reste encore  $j-1$  éléments à placer, donc en particulier le nombre total de cycles d'une permutation aléatoire  $\sigma$  de  $[n]$  est  $K(\sigma) = \xi_1 + \dots + \xi_n$ . Les nombres de cycles de mêmes tailles s'expriment aussi à l'aide des  $\xi_j$ , et on peut montrer que les limites lorsque  $n$  tend vers l'infini de leurs expressions forment des variables de Poisson indépendantes de paramètres  $\frac{\theta}{j}$  (voir [4] pour plus de détails). Une idée clé est alors d'utiliser le couplage de Feller pour

- remplacer dans les expressions qu'on manipule tous les nombres de cycles de mêmes tailles par leurs variables de Poisson correspondantes.
- estimer l'erreur qu'on fait en faisant ces remplacements.

On se ramène ainsi à des expressions dont les moments sont plus faciles à calculer, d'une part car la loi de Poisson est beaucoup plus simple que celle des nombres de cycles de

mêmes tailles, et d'autre part car les variables de Poisson obtenues par le couplage de Feller sont indépendantes, ce qui n'est pas le cas des nombres de cycles de mêmes tailles (contraints par la relation  $\sum_{j=1}^n j a_j = n$ ).

Une autre manière bien connue pour générer des permutations de loi d'Ewens est le *processus des restaurants chinois* (voir par exemple [27]), que l'on peut décrire comme suit : on considère un restaurant initialement sans client avec une infinité de tables rondes, chacune de capacité infinie. Les clients arrivent un par un et s'installent à une table selon cette règle : le client 1 s'installe à une table, puis, une fois que  $n \geq 1$  clients sont déjà installés, le client  $n + 1$  s'installe à une table non vide à un des  $n$  emplacements possibles entre les clients déjà assis avec probabilité  $\frac{1}{\theta+n}$  pour chacun de ces emplacements, et s'installe à une nouvelle table avec probabilité  $\frac{\theta}{\theta+n}$ . Pour tout  $m \leq n$ , la lecture dans le sens trigonométrique des numéros de clients  $\leq m$  de chaque table donne des cycles, et ces cycles constituent une unique permutation  $\sigma_m$  de  $[m]$ . En faisant s'installer les clients jusqu'à l'infini, la suite de permutations  $(\sigma_m)_{m \geq 1}$  obtenue est une *permutation virtuelle*. Autrement dit,  $\sigma_m$  s'obtient à partir de  $\sigma_{m+1}$  en retirant simplement l'élément  $m + 1$  de son écriture en cycles à support disjoint. On dit aussi que  $\sigma_m$  est la *permutation dérivée* de  $\sigma_{m+1}$ .

Cette méthode est en quelque sorte duale de celle de Feller : au lieu de considérer les tables une par une et de les remplir jusqu'à un certain temps en choisissant au hasard les clients, on part des clients un par un et on les installe à des tables. Elle présente aussi des avantages différents, comme par exemple le fait d'être une construction markovienne dont la connaissance du présent détient l'information du passé dans son intégralité.

### 1.1.3 Lois plus générales

Les lois d'Ewens peuvent être naturellement généralisées en mettant des poids différents sur chaque taille de cycles, c'est-à-dire en attribuant à une permutation de  $[n]$  de structure en cycles  $(a_1, \dots, a_n)$  le poids

$$\frac{1}{n! h_n} \prod_{j=1}^n \theta_j^{a_j}$$

où  $h_n$  est une constante de normalisation. Les lois d'Ewens classiques correspondent au cas  $\theta_k = \theta$  constant. Ces généralisations de lois d'Ewens ont été beaucoup étudiées en 2010-13 par Nikeghbali et Zeindler [73], puis par Hugues, Najnudel, Nikeghbali et Zeindler [52]. Un des inconvénients est que la méthode de couplage de Feller ne s'étend pas, autrement dit il n'y a pas de couplage analogue qu'on pourrait prendre pour généraliser facilement les résultats. Les outils utilisés dans les preuves sont d'ordre combinatoires et analytiques, avec des manipulations de fonctions génératrices.

De manière plus générale, les groupes symétriques peuvent être munis de *mesures centrales*, c'est-à-dire des mesures constantes sur chaque classe de conjugaison. En particulier les lois d'Ewens généralisées sont des mesures centrales.

Kerov, Olshanski et Vershik [60] [58] introduisent la notion de *permutations virtuelles* et étendent le concept de mesures centrales à des groupes infinis de permutations en utilisant des résultats analytiques et de théorie des représentations de groupes.

L'avantage de considérer des mesures centrales est que cela se rapporte à des mesures sur des partitions d'entiers, en ce fait que les partitions de l'entier  $n$  paramétrisent les classes de conjugaisons de  $\mathfrak{S}_n$ . De plus, comme nous l'expliquons dans la section suivante, le spectre d'une matrice de permutation est entièrement caractérisé par la structure en cycles de la permutation associée (et réciproquement). Par conséquent, la mesure spectrale empirique d'une matrice de permutation de taille  $n$  se rapporte à une mesure empirique sur les partitions de l'entier  $n$ .

## 1.2 Matrices de permutation

Les matrices de permutation sont des matrices associées aux endomorphismes qui permutent les vecteurs de base. Plus précisément, si  $(e_1, \dots, e_n)$  est la base canonique de  $\mathbb{C}^n$ , et  $\sigma$  est une permutation de  $[n]$ , on appelle matrice de permutation associée à  $\sigma$  la matrice  $M_\sigma$  de taille  $n \times n$  telle que pour tout  $i$ ,  $M_\sigma e_i = e_{\sigma(i)}$ . Les entrées de  $M_\sigma$  sont

$$\text{alors données par } (M_\sigma)_{i,j} = \begin{cases} 1 & \text{si } i = \sigma(j) \\ 0 & \text{sinon.} \end{cases}.$$

On appelle  $S_n := \{M_\sigma : \sigma \in \mathfrak{S}_n\}$  l'ensemble des **matrices de permutation** de taille  $n$ . Une manière de décrire l'ensemble des matrices de permutation est de dire qu'il est l'ensemble des matrices où sur chaque colonne et chaque ligne il y a exactement un coefficient égal à 1 et que des 0 ailleurs. On peut aussi l'exprimer comme étant l'ensemble des matrices bistochastiques à coefficients entiers.

Une propriété élémentaire est que l'application  $\sigma \mapsto M_\sigma$  définit un isomorphisme de groupes entre  $\mathfrak{S}_n$  et  $S_n$  :

$$\forall \sigma, \tau \in \mathfrak{S}_n, \quad M_\sigma \cdot M_\tau = M_{\sigma \circ \tau}$$

ce qui fait de  $S_n$  un sous-groupe du groupe des matrices unitaires  $U_n$ . En particulier, l'inclusion de  $S_n$  dans  $U_n$  implique que les valeurs propres de matrices de permutation appartiennent au cercle unité.

### 1.2.1 Lien entre valeurs propres et structure en cycles

La donnée de la structure en cycles d'une permutation renseigne totalement sur le spectre de sa matrice de permutation associée. Nous présentons ci-après une manière très simple de le voir et d'en tirer une formulation explicite.

Soit  $\sigma \in \mathfrak{S}_n$ , et notons  $(a_1, \dots, a_n)$  sa structure en cycles. Écrivons  $\sigma$  en cycles à supports disjoints dans un ordre croissant de tailles de cycles (d'abord les points fixes, puis les cycles de taille 2, etc...),

$$\begin{aligned} \sigma = & (w_1) \dots (w_{a_1}) (w_{a_1+1} \ w_{a_1+2}) \dots (w_{a_1+2a_2-1} \ w_{a_1+2a_2}) \\ & \dots (w_{a_1+2a_2+\dots+ka_k-k+1} \ \dots \ w_{a_1+2a_2+\dots+ka_k}), \end{aligned}$$

avec  $k$  la plus grande taille de cycle de  $\sigma$ . Soit  $N_\sigma$  la matrice carrée de taille  $n$  définie comme  $(N_\sigma)_{\tau(i), \tau(j)} := (M_\sigma)_{i,j}$ , où  $\tau$  est la permutation de  $[n]$  envoyant chaque  $w_j$  sur  $j$ . En d'autres termes,  $N_\sigma$  est la matrice obtenue à partir de  $M_\sigma$  en permutant



simultanément ses colonnes et ses lignes selon  $\tau$ . On a

$$(N_\sigma)_{i,j} = (M_\sigma)_{\tau^{-1}(i),\tau^{-1}(j)} = \delta_{\tau^{-1}(i),\sigma(\tau^{-1}(j))} = \delta_{i,(\tau\sigma\tau^{-1})(j)} = (M_{\tau\sigma\tau^{-1}})_{i,j}$$

où  $\delta_{i,j}$  est le symbole de Kronecker. Autrement dit  $N_\sigma$  est la matrice de permutation de  $\tau\sigma\tau^{-1}$ , qui a la même structure en cycle que  $\sigma$ . De plus, comme  $M_{\tau\sigma\tau^{-1}} = M_\tau M_\sigma (M_\tau)^{-1}$  par morphisme de groupes, alors  $N_\sigma$  et  $M_\sigma$  sont des matrices semblables et donc ont même polynôme caractéristique. Au passage, on déduit que le spectre des matrices de permutation est constant sur chaque classe de conjugaison des permutations, et donc ne dépend que de la structure en cycle.

La Figure 1 présente la forme de la matrice  $N_\sigma$ . Elle est diagonale par blocs, dont chaque bloc de taille  $r$  (correspondant à un cycle de taille  $r$  dans la permutation  $\tau\sigma\tau^{-1}$ ) est la matrice compagnon du polynôme  $X^r - 1$ . Ainsi, chaque  $r$ -cycle de  $\sigma$  correspond à un ensemble de  $r$  valeurs propres :  $\{1, e^{2i\pi/r}, e^{4i\pi/r}, \dots, e^{2i(r-1)\pi/r}\}$ , et donc pour tout  $r$ ,  $M_\sigma$  possède  $a_r$  copies de ces valeurs propres. Reformulé avec le polynôme caractéristique,

$$\det(XI - M_\sigma) = \prod_{r=1}^n (X^r - 1)^{a_r}.$$

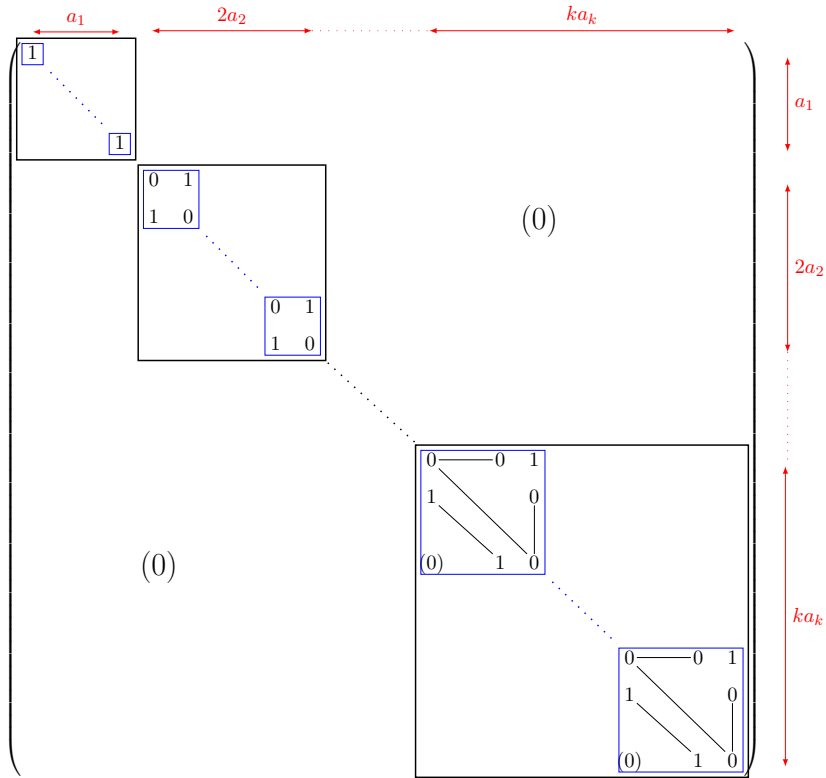


Figure 1 – Matrice  $N_\sigma$ , où  $k$  est la taille du plus grand cycle de  $\sigma$ .

En résumé, nous venons de voir que les valeurs propres de matrices de permutation sont des racines de l'unité dont les nombres d'exemplaires sont directement déterminés

par les structures en cycle des permutations associées. Une question naturelle est alors de se demander comment se répartissent ces valeurs propres sur le cercle unité lorsqu'on tire les permutations selon une loi donnée sur  $\mathfrak{S}_n$  et qu'on fait tendre  $n$  vers l'infini.

### 1.2.2 Répartition des valeurs propres

Une des motivations principales pour l'étude des propriétés spectrales de matrices de permutation aléatoires est de comparer les résultats obtenus avec d'autres ensembles de matrices aléatoires. Notamment, en quelle mesure certaines propriétés des groupes de permutations (finis) se retrouvent dans celles de groupes compacts plus grands tels que le groupe orthogonal ou le groupe unitaire (infinis).

Une manière intuitive, développée par Wieand, d'étudier la répartition des valeurs propres est de se fixer des portions du cercle unité et de compter les valeurs propres tombant dedans.

Notons  $I := ]e^{2i\pi\alpha}, e^{2i\pi\beta}]$  l'arc du cercle unité parcouru dans le sens trigonométrique du point  $e^{2i\pi\alpha}$  (exclus) au point  $e^{2i\pi\beta}$  (inclus), où  $0 \leq \alpha < 1$  et  $\alpha \leq \beta < \alpha + 1$  (ces conditions sur  $\alpha$  et  $\beta$  suffisent à pouvoir prendre l'intervalle semi-ouvert que l'on veut du cercle unité).

**Théorème 1.2.1** (Wieand, 1999 [87]). *Pour  $n \geq 1$  posons  $X_n^I$  la variable aléatoire représentant le nombre de valeurs propres dans  $I$  d'une matrice de permutation aléatoire, dont la permutation associée suit la loi uniforme sur  $\mathfrak{S}_n$ . Alors*

$$\begin{cases} \mathbb{E}(X_n^I) & \underset{n \rightarrow \infty}{=} n(\beta - \alpha) + c_1 \ln n + o(\ln n) \\ \text{Var}(X_n^I) & \underset{n \rightarrow \infty}{=} c_2 \ln n + o(\ln n) \end{cases}$$

où  $c_1$  et  $c_2$  sont des constantes dépendant uniquement de  $\alpha$  et  $\beta$ . De plus, les fluctuations de  $X_n^I$  sont asymptotiquement gaussiennes, c'est-à-dire qu'on a la convergence en loi

$$Y_n^I := \frac{X_n^I - \mathbb{E}(X_n^I)}{\sqrt{\text{Var}(X_n^I)}} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1).$$

Plus généralement, pour tout nombre fini fixé  $m \in \mathbb{N}^*$  d'arcs  $I_1, \dots, I_m$ , on a la convergence en loi

$$(Y_n^{I_1}, \dots, Y_n^{I_m}) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, D)$$

où  $D$  est une matrice de covariance dont on sait calculer les coefficients.

### 1.2.3 Statistiques linéaires

En 2011, Ben Arous et Dang s'intéressent à des statistiques linéaires du spectre des matrices de permutation sous des mesures d'Ewens, et établissent les résultats suivants : Pour  $N \geq 1$ , considérons  $\sigma$  une permutation de taille  $N$  et écrivons les valeurs propres de sa matrice associée

$$\lambda_1(\sigma) = e^{2i\pi\varphi_1(\sigma)}, \dots, \lambda_N(\sigma) = e^{2i\pi\varphi_N(\sigma)},$$

où  $\varphi_1(\sigma), \dots, \varphi_N(\sigma) \in [0, 1[$ . Pour toute fonction 1-périodique  $f : \mathbb{R} \rightarrow \mathbb{R}$  on définit

$$I_{\sigma,N}(f) := \sum_{i=1}^N f(\varphi_i(\sigma))$$

et pour tout  $j \geq 1$ ,

$$R_j(f) := \frac{1}{j} \sum_{k=0}^{j-1} f\left(\frac{k}{j}\right) - \int_0^1 f(x) dx.$$

Les  $R_j(f)$  s'interprètent comme des termes d'erreurs dans l'approximation de l'intégrale de  $f$  sur  $[0, 1]$  par la méthode des trapèzes.

**Théorème 1.2.2** (Ben Arous, Dang, 2011 [13]). *Soit  $\theta > 0$ . Soit  $f$  une fonction à variation bornée telle que*

$$\sum_{j=1}^{+\infty} j R_j(f)^2 \in ]0, +\infty[.$$

*Alors sous  $E_{\text{wens}}(\theta)$ ,  $I_{\sigma,N}(f) - \mathbb{E}(I_{\sigma,N}(f))$  converge en distribution vers une loi infinidivisible non gaussienne  $\mu_{f,\theta}$  dont la transformée de Fourier est donnée par*

$$\widehat{\mu}_{f,\theta}(t) = \exp\left(\theta \int (e^{itx} - 1 - itx) dM_f(x)\right)$$

où  $M_f = \sum_{j=1}^{+\infty} \frac{1}{j} \delta_{jR_j(f)}$ . De plus,

$$\mathbb{E}(I_{\sigma,N}(f)) = N \int_0^1 f(x) dx + \theta \sum_{j=1}^N R_j(f) + o(1)$$

et

$$\text{Var}(I_{\sigma,N}(f)) = \theta \sum_{j=1}^N j R_j(f)^2 + o(1)$$

avec  $\sum_{j=1}^N R_j(f) = \mathcal{O}(\sqrt{\ln N})$ .

**Théorème 1.2.3** (Ben Arous, Dang, 2011 [13]). *Soit  $\theta > 0$ . Soit  $f$  une fonction à variation bornée telle que*

$$\sum_{j=1}^{+\infty} j R_j(f)^2 = +\infty.$$

*Alors sous  $E_{\text{wens}}(\theta)$ , on a la convergence en distribution*

$$\frac{I_{\sigma,N}(f) - \mathbb{E}(I_{\sigma,N}(f))}{\sqrt{\text{Var}(I_{\sigma,N}(f))}} \longrightarrow \mathcal{N}(0, 1).$$

De plus,

$$\mathbb{E}(I_{\sigma,N}(f)) = N \int_0^1 f(x) dx + \theta \sum_{j=1}^N R_j(f) + \mathcal{O}(1)$$

et

$$\mathrm{Var}(I_{\sigma,N}(f)) \sim \theta \sum_{j=1}^N j R_j(f)^2$$

avec  $\sum_{j=1}^N R_j(f) = \mathcal{O}(\ln N)$ .

En d'autres termes, pour des fonctions suffisamment régulières (premier théorème) la variance ne diverge pas et donc il n'y a pas besoin de normalisation pour obtenir une loi limite, et en revanche pour des fonctions moins régulières (deuxième théorème) la variance diverge lentement et on a une loi limite gaussienne. Le Théorème 1.2.1 dans le cas d'un arc est un corollaire du Théorème 1.2.3 en prenant pour  $f$  l'indicatrice  $\mathbb{1}_{] \alpha, \beta]}$  périodisée, et  $\theta = 1$ .

#### 1.2.4 Polynômes caractéristiques et fonctions centrales

Les polynômes caractéristiques contiennent toute l'information du spectre, et sont l'objet de beaucoup d'attention en théorie des matrices aléatoires, particulièrement ces dernières années depuis qu'a été révélé un lien avec les zéros non triviaux de la fonction zeta de Riemann (voir Keating et Snaith [57]).

Concernant les matrices de permutation aléatoires, Hambly, Keevash, O'Connell, et Stark établissent le résultat suivant :

Pour  $M \in \mathcal{M}_n(\mathbb{C})$  et  $x \in \mathbb{C}$  on pose

$$L_M(x) := \begin{cases} \log(\det(I - xM)) & \text{si cela est bien défini} \\ 0 & \text{sinon} \end{cases},$$

où  $\log$  est l'unique détermination du logarithme continue sur l'ouvert connexe  $\mathbb{C}$  privé des demi-droites des valeurs propres inverses vers l'infini, et valant 0 en 1.

**Théorème 1.2.4** (Hambly, Keevash, O'Connell, Stark, 2000 [48]). *Soit  $M_n$  une matrice de permutation aléatoire de taille  $n$ , tirée selon la loi uniforme. Soit  $\alpha \in [0, 1[$  irrationnel. Alors :*

1. *Lorsque  $n$  tend vers l'infini on a la convergence en loi*

$$\Im \left( \frac{L_{M_n}(e^{2i\pi\alpha}) - \mathbb{E}L_{M_n}(e^{2i\pi\alpha})}{\sqrt{\frac{\pi^2}{12} \ln n}} \right) \rightarrow \mathcal{N}(0, 1).$$

2. *De plus, si  $\alpha$  est de type fini (voir définition ci-après), alors les parties réelle et imaginaire de  $\frac{L_{M_n}(e^{2i\pi\alpha})}{\sqrt{\frac{\pi^2}{12} \ln n}}$  convergent chacune en loi vers une  $\mathcal{N}(0, 1)$ .*

Le type d'un nombre réel traduit le comportement de la distance au plus proche entier (notée  $\|.\|$ ) de ses multiples. En voici la définition.

**Définition 1.2.5.** Soit  $x \in \mathbb{R}$ . Posons

$$\eta = \sup\{\gamma \in \mathbb{R} : \liminf_{n \rightarrow +\infty} n^\gamma \|nx\| = 0\} \in \mathbb{R}_+ \cup \{+\infty\}.$$

On dit que  $x$  est de **type fini** lorsque  $\eta$  est fini.

Zeindler [93] va plus loin en montrant que lorsque  $\alpha$  est de type fini, les deux variables gaussiennes vers lesquelles tendent les parties réelle et imaginaire de  $\frac{L_{Mn}(e^{2i\pi\alpha})}{\sqrt{\frac{\pi^2}{12} \ln n}}$  sont indépendantes. Il étend même le résultat précédent à des *fonctions centrales* (la fonction polynôme caractéristique en est une en particulier) et pour des permutations tirées sous Ewens( $\theta$ ),  $\theta > 0$ . Plus précisément :

**Définition 1.2.6.** Soit  $x \in S^1$  et soit  $f : S^1 \rightarrow \mathbb{C}$  une fonction régulière. On appelle fonction centrale multiplicative associée à  $f$

$$W^{(n)}(f) = W^{(n)}(f)(x)(\sigma) := \prod_{j=1}^n f(x^j)^{a_{n,j}}$$

où on rappelle que  $a_{n,j}$  désigne le nombre de cycles de taille  $j$  dans une permutation de  $[n]$ . On définit aussi

$$w^{(n)}(f)(x) = \log(W^{(n)}(f)(x)) = \sum_{j=1}^n a_{n,j} \log f(x^j).$$

Il est clair que  $W^{(n)}(f)$  et  $w^{(n)}(f)$  sont des fonctions centrales, et que le polynôme caractéristique, avec la convention prise ici, est le cas particulier où  $f(x)$  vaut  $1 - x$ .

**Définition 1.2.7.** Soit  $f : S^1 \rightarrow \mathbb{C}$  une fonction régulière et  $x \in S^1$ . Si  $x$  est une racine  $p$ -ième primitive de l'unité, alors on définit

$$m(f)(x) := \frac{1}{p} \sum_{j=1}^p \log f(x^j)$$

et si  $x$  n'est pas une racine de l'unité,

$$m(f)(x) := \int_0^1 \log f(e^{2i\pi s}) ds.$$

Dans le cas où la somme ou l'intégrale n'existe pas on pose  $m(f)(x) = +\infty$ .

**Théorème 1.2.8** (Zeindler, 2011 [93]). Soit  $\theta > 0$ . Soit  $f : S^1 \rightarrow \mathbb{C}$  une fonction régulière et  $x \in S^1$ . Alors, sous Ewens( $\theta$ ),

- Si  $x$  n'est pas une racine de l'unité et est de type fini, et que tous les zéros de  $f$  sont des racines de l'unité, alors on a la convergence en loi lorsque  $n$  tend vers l'infini

$$\frac{w^{(n)}(f)(x)}{\sqrt{\ln n}} - \theta m(f)(x) \sqrt{\ln n} \longrightarrow \mathcal{N}$$

où  $\mathcal{N}$  est une variable complexe gaussienne. De plus, dans ce cas la covariance des parties réelle et imaginaire de  $\mathcal{N}$  est donnée par

$$\frac{\theta}{2} \Im \left( \int_0^1 \log^2 f(e^{2i\pi s}) ds \right).$$

- Si  $x$  est une racine de l'unité d'ordre  $p$  et  $f(x^j) \neq 0$  pour tous  $1 \leq m \leq p$ , alors

$$\frac{w^{(n)}(f)(x)}{\sqrt{\ln n}} - \theta m(f)(x) \sqrt{\ln n} \longrightarrow \mathcal{N}$$

où  $\mathcal{N}$  est une variable complexe gaussienne. De plus, dans ce cas la covariance des parties réelle et imaginaire de  $\mathcal{N}$  est donnée par

$$\frac{\theta}{2} \Im \left( \frac{1}{p} \sum_{j=1}^p \log^2 f(x^j) \right).$$

Dang et Zeindler [34] généralisent ce résultat en considérant plusieurs points au lieu d'un seul. Ils obtiennent le théorème suivant.

**Théorème 1.2.9** (Dang, Zeindler, 2013 [34]). *Soit  $\theta > 0$ . Soient  $f_1, \dots, f_d : S^1 \rightarrow \mathbb{C}$  des fonctions régulières et  $x_1 = e^{2i\pi\varphi_1}, \dots, x_d = e^{2i\pi\varphi_d} \in S^1$  tels que  $1, \varphi_1, \dots, \varphi_d$  sont  $\mathbb{Z}$ -libres. Supposons que tous les zéros des  $f_j$  sont des racines de l'unité. Supposons de plus que  $\varphi_1, \dots, \varphi_d$  sont deux à deux de type fini, c'est-à-dire pour tous  $j \neq k$  il existe  $K > 0$  et  $\gamma \geq 1$  tels que*

$$\forall (m_1, m_2) \in \mathbb{Z}^2 \setminus \{0\}, \quad \|m_1\varphi_j + m_2\varphi_k\| \geq \frac{K}{\max(|m_1|, |m_2|)^\gamma}. \quad (1.1)$$

Alors, sous  $E_{\text{wens}}(\theta)$ , lorsque  $n$  tend vers l'infini on a la convergence en loi

$$\frac{1}{\sqrt{\ln n}} \begin{pmatrix} w^{(n)}(f_1)(x_1) \\ \vdots \\ w^{(n)}(f_d)(x_d) \end{pmatrix} - \theta \sqrt{\ln n} \begin{pmatrix} m(f_1) \\ \vdots \\ m(f_d) \end{pmatrix} \longrightarrow N = \begin{pmatrix} N_1 \\ \vdots \\ N_d \end{pmatrix}$$

où  $N$  est une variable gaussienne complexe  $d$ -dimensionnelle telle que pour  $j \neq k$ ,

$$\begin{aligned} \text{Cov}(\Re(N_j), \Re(N_k)) &= \theta \int_{[0,1]^2} \ln |f_j(e^{2i\pi u})| \ln |f_k(e^{2i\pi v})| du dv, \\ \text{Cov}(\Re(N_j), \Im(N_k)) &= \theta \int_{[0,1]^2} \ln |f_j(e^{2i\pi u})| \arg(f_k(e^{2i\pi v})) du dv, \\ \text{Cov}(\Im(N_j), \Im(N_k)) &= \theta \int_{[0,1]^2} \arg(f_j(e^{2i\pi u})) \arg(f_k(e^{2i\pi v})) du dv, \end{aligned}$$

et pour  $j = k$ ,

$$\begin{aligned} \text{Var}(\Re(N_j)) &= \theta \int_0^1 \log^2 |f_j(e^{2i\pi u})| du, \\ \text{Var}(\Im(N_j)) &= \theta \int_0^1 \arg^2(f_j(e^{2i\pi v})) dv. \end{aligned}$$

Un corollaire immédiat en découle pour le polynôme caractéristique  $Z_n(x) := \det(I - x^{-1}M_n)$  :

**Corollaire 1.2.10** (Dang, Zeindler, 2013 [34]). Soient  $x_1 = e^{2i\pi\varphi_1}, \dots, x_d = e^{2i\pi\varphi_d} \in S^1$  tels que  $\varphi_1, \dots, \varphi_d$  sont deux à deux de type fini. Alors, sous  $Ewens(\theta)$ , lorsque  $n$  tend vers l'infini on a la convergence en loi

$$\frac{1}{\sqrt{\frac{\pi^2}{12}\theta \ln n}} \begin{pmatrix} \log(Z_n(x_1)) \\ \vdots \\ \log(Z_n(x_d)) \end{pmatrix} \rightarrow \begin{pmatrix} N_1 \\ \vdots \\ N_d \end{pmatrix}$$

où  $\Re(N_1), \Im(N_1), \dots, \Re(N_d), \Im(N_d)$  sont des variables gaussiennes réelles indépendantes centrées réduites.

### 1.3 Matrices de permutation modifiées

Comme nous l'avons dit précédemment, une des motivations à considérer les permutations en tant que matrices est de pouvoir analyser et comparer les propriétés de ces matrices à d'autres ensembles de matrices aléatoires. Beaucoup d'autres groupes que le groupe symétrique peuvent s'identifier à des groupes de matrices dont les éléments ont sur chaque ligne et chaque colonne exactement une entrée non nulle et que des zéros ailleurs. C'est le cas notamment de certains *produits en couronne* avec le groupe symétrique. Sans entrer dans les détails avec la définition abstraite du produit en couronne en théorie des groupes, mentionnons simplement que pour deux groupes  $G$  et  $F$  celui-ci se note  $G \wr F$ , et que par exemple  $\mathfrak{S}_n \wr \mathfrak{S}_n$  s'identifie aux matrices de permutation par blocs où les blocs sont eux-mêmes des matrices de permutation. Avec la même logique, on peut voir que l'ensemble des automorphismes sur les arbres  $n$ -réguliers de taille fixée  $d$  s'identifie au produit en couronne de  $\mathfrak{S}_n$  avec lui-même  $d$  fois  $\mathfrak{S}_n \wr \dots \wr \mathfrak{S}_n$  (voir Evans [39]). Ce qui est intéressant avec ces groupes est qu'ils se ramènent à des groupes matriciels dont les éléments  $M$  s'écrivent sous la forme

$$M = (M_{i,j})_{1 \leq i,j \leq N} = (b_i \mathbb{1}_{i=\sigma(j)})_{1 \leq i,j \leq N}$$

où  $\sigma \in \mathfrak{S}_N$  et  $b_1, \dots, b_N$  sont des nombres, pour un certain  $N$  qui est le même pour tous les éléments  $M$ .

Najnudel et Nikeghbali [68] étudient des matrices de permutation aléatoires où les entrées égales à 1 sont remplacées par des variables aléatoires à valeurs dans  $\mathbb{C}^*$ , indépendantes et de même loi. Les réalisations de ces matrices aléatoires s'identifient donc à des éléments de  $\mathbb{C}^* \wr \mathfrak{S}_n$ , où  $n$  est la taille des matrices. En munissant  $\mathfrak{S}_n$  d'une loi d'Ewens, Najnudel et Nikeghbali obtiennent des résultats pour des entrées non nulles de lois générales  $\mathcal{L}$  sur  $\mathbb{C}^*$ , avec certaines hypothèses sur  $\mathcal{L}$ .

Dans cette thèse, le cas qui nous intéresse est le cas particulier où les variables aléatoires qui remplacent les 1 sont indépendantes et uniformes sur le cercle unité. Les matrices de permutation obtenues par cette modification restent unitaires. Dans toute la suite, nous appellerons **matrices de permutation modifiées** les matrices aléatoires de cette forme. Autrement dit, si  $M_n$  désigne la matrice de permutation aléatoire

$$M_n = ((M_n)_{i,j})_{1 \leq i,j \leq n} = (\mathbb{1}_{i=\sigma_n(j)})_{1 \leq i,j \leq n},$$

où  $\sigma_n$  est une permutation aléatoire de loi d'Ewens( $\theta$ ) sur  $\mathfrak{S}_n$ , alors la matrice de permutation modifiée  $\widetilde{M}_n$  associée à  $\sigma_n$  s'écrit

$$\widetilde{M}_n = ((\widetilde{M}_n)_{i,j})_{1 \leq i,j \leq n} = (z_i \mathbb{1}_{i=\sigma_n(j)})_{1 \leq i,j \leq n},$$

où  $z_1, \dots, z_n$  sont des variables aléatoires indépendantes et uniformément distribuées sur  $S^1$ .

Un des aspects avantageux des matrices de permutation modifiées est que leur spectre est simple à décrire par rapport aux permutations associées. En effet, étant donnée une matrice de permutation, la modification entraîne que pour chaque  $j$ -cycle de la permutation associée, les angles propres associés de la forme  $\frac{2\pi k}{j}$  sont tous translatés d'une même variable aléatoire  $\varphi$  uniforme sur  $[0, 2\pi[$ . De plus, les variables  $\varphi$  relatives à chaque  $j$ -cycle sont indépendantes.

### 1.3.1 Similarités avec le CUE

L'ensemble circulaire unitaire (Circular Unitary Ensemble en anglais), est l'ensemble des matrices aléatoires à valeurs dans un groupe unitaire  $U(n)$  sous la mesure de Haar. Autrement dit, pour tout élément  $U$  de cet ensemble, la loi de  $U$  est invariante par multiplication à gauche ou à droite par une matrice unitaire déterministe de même taille. Étant donné un élément  $U$  de taille  $n$ , la loi jointe des valeurs propres  $e^{i\theta_1}, \dots, e^{i\theta_n}$  admet une densité par rapport à la mesure uniforme sur  $(S^1)^n$ , donnée par

$$\frac{1}{h_n} \prod_{1 \leq j < k \leq n} |e^{i\theta_j} - e^{i\theta_k}|^2$$

où  $h_n$  est une constante de renormalisation. Les valeurs propres ont donc tendance à se repousser les unes des autres sur le cercle unité. De plus, on a l'écriture déterminantale

$$\prod_{1 \leq j < k \leq n} |e^{i\theta_j} - e^{i\theta_k}|^2 = \det((K_n(\theta_j, \theta_k))_{1 \leq j, k \leq n})$$

où

$$K_n(x, y) := \frac{\sin(n(x - y)/2)}{\sin((x - y)/2)}.$$

Nous renvoyons au livre de Mehta [66] pour de nombreux résultats classiques sur les ensembles de matrices unitairement invariant et des applications en physique.

Des liens avec les permutations apparaissent dans certaines études du CUE. Nous donnons brièvement ci-après quelques résultats illustrant ce propos.

En 1994, Diaconis [36] s'intéresse aux traces des puissances successives d'un élément  $U$  du CUE et pour l'espérance du produit en dégage une écriture à l'aide de caractères irréductibles associés à des groupes symétriques. En particulier pour  $U$  de taille  $k$  et tout  $n \leq k$  il obtient

$$\mathbb{E}(|\text{Tr}(U)^n|^2) = n!.$$

En 1998, Rains [79], par rapport à ce résultat particulier de Diaconis se demande ce qu'il



en est pour  $n > k$ , et découvre que

$$f_{nk} := \mathbb{E}(|\text{Tr}(U)^n|^2)$$

est le nombre de permutations de  $[n]$  n'ayant pas de sous-suite croissante de longueur strictement plus grande que  $k$ .

En 1999, Tracy et Widom [84], suite notamment à ces travaux et à l'article percutant de Baik, Deift et Johansson [8], étudient la fonction caractéristique de  $\text{Tr}(U)$  et soulignent des liens avec plusieurs équations de type Painlevé et les plus longues sous-suites croissantes de permutations.

Les matrices de permutation forment un sous-groupe fini du groupe unitaire, tandis que les matrices de permutation modifiées forment un sous-groupe infini.

Tout comme le CUE, pour tout arc  $I$  du cercle unité, l'espérance du nombre de valeurs propres dans  $I$  des matrices de permutation modifiées est proportionnel à la longueur de  $I$ . Cependant, à la différence du CUE, le processus des angles propres des matrices de permutation modifiées n'est pas déterminantal. Pour l'étude du spectre on exploite la structure discrète des permutations.

Dans la suite, nous allons voir que les résultats présentés pour les matrices de permutation ont des versions très analogues pour les matrices de permutation modifiées.

### 1.3.2 Répartition des valeurs propres

Dans [89], Wieand étudie des matrices aléatoires reliées à des permutations, où les permutations sont tirées sous la loi uniforme. Les ensembles auxquels elle s'intéresse s'identifient à des produits en couronne de deux groupes dont le deuxième est le groupe symétrique. Plus spécifiquement, Wieand étudie les produits en couronnes de la forme  $(\mathbb{Z}_{K_1} \times \cdots \times \mathbb{Z}_{K_m}) \wr \mathfrak{S}_n$  et  $\mathbb{T}^m \wr \mathfrak{S}_n$ , où  $\mathbb{Z}_K$  est le groupe des racines  $K$ -ème de l'unité et  $\mathbb{T} = S^1$  est le groupe des complexes de module un. Pour chacun de ces ensembles elle montre que sous la mesure uniforme les fluctuations des nombres de valeurs propres appartenant à des arcs donnés fixés du cercle unité sont asymptotiquement gaussiennes, et en détermine la matrice de covariance limite. Elle compare la structure de covariance trouvée à celle pour le groupe des matrices unitaires sous la loi de Haar [88]. Nous présentons ici de manière simplifiée son résultat concernant le cas particulier qui nous intéresse dans cette thèse, à savoir  $\mathbb{T}^m \wr \mathfrak{S}_n$  avec  $m = 1$ .

**Théorème 1.3.1** (Wieand, 2003 [89]). *Pour  $n \geq 1$  posons  $\tilde{X}_n^I$  la variable aléatoire représentant le nombre de valeurs propres dans  $I := ]e^{2i\pi\alpha}, e^{2i\pi\beta}]$  d'un élément de  $S^1 \wr \mathfrak{S}$  suivant la loi uniforme. Alors*

$$\begin{cases} \mathbb{E}(\tilde{X}_n^I) &= n(\beta - \alpha) \\ \text{Var}(\tilde{X}_n^I) &\underset{n \rightarrow \infty}{=} \ell \ln n + o(\ln n) \end{cases}$$

où  $\ell$  est une constante dépendant uniquement de  $\alpha$  et  $\beta$ . De plus, les fluctuations de  $\tilde{X}_n^I$

sont asymptotiquement gaussiennes, c'est-à-dire qu'on a la convergence en loi

$$\tilde{Y}_n^I := \frac{\tilde{X}_n^I - \mathbb{E}(\tilde{X}_n^I)}{\sqrt{\text{Var}(\tilde{X}_n^I)}} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1).$$

Plus généralement, pour tout nombre fini fixé  $m \in \mathbb{N}^*$  d'arcs  $I_1, \dots, I_m$ , on a la convergence en loi

$$(\tilde{Y}_n^{I_1}, \dots, \tilde{Y}_n^{I_m}) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \tilde{D})$$

où  $\tilde{D}$  est une matrice de covariance dont on sait calculer les coefficients.

Sans entrer dans les détails, la structure de covariance ici n'est pas simple et dépend des propriétés arithmétiques des extrémités des arcs  $I_1, \dots, I_m$ . Pour le CUE en revanche, la structure de covariance est beaucoup plus simple et frappante : en notant  $e^{2i\pi\alpha_j}, e^{2i\pi\beta_j}$  les extrémités de  $I_j$ , et  $Z_j$  le nombre de valeurs propres appartenant à  $I_j$  d'un élément du CUE, on a

$$\text{Cov}(Z_1, Z_2) = \begin{cases} 0 & \text{si } \alpha_1, \alpha_2, \beta_1, \beta_2 \text{ sont deux à deux distincts} \\ 1/2 & \text{si } \alpha_1 = \alpha_2 \text{ ou } \beta_1 = \beta_2 \\ -1/2 & \text{si } \alpha_1 = \beta_2 \text{ ou } \beta_1 = \alpha_2 \end{cases}.$$

Même si ces deux structures de covariance diffèrent, Wieand remarque qu'elles sont en fait identiques pour presque tous les arcs, y compris ceux partageant une même extrémité. En effet, en prenant pour simplifier le cas de deux arcs  $I_1$  et  $I_2$ , si les quantités non-nulles parmi  $\alpha_1 - \alpha_2, \beta_1 - \beta_2, \alpha_1 - \beta_2$  et  $\beta_1 - \alpha_2$  sont des irrationnels, alors  $\tilde{D}_{1,2} = \text{Cov}(Z_1, Z_2)$ . En imaginant qu'on tire les extrémités des arcs  $I_1$  et  $I_2$  de manière uniforme sur le cercle, cela arrive presque sûrement.

### 1.3.3 Statistiques linéaires

Dang, dans sa thèse [33], poursuit l'étude réalisée avec Ben Arous sur des statistiques linéaires du spectre des matrices de permutation sous des mesures d'Ewens, en considérant des modifications de matrices de permutation où les entrées égales 1 sont remplacées par des variables  $z_i$  i.i.d à valeurs dans le cercle unité (pas forcément de loi uniforme). Elle obtient des résultats très analogues, que voici.

Pour  $N \geq 1$ , considérons  $\sigma$  une permutation de taille  $N$ , et

$$M_{\sigma,z} := (z_i \mathbf{1}_{i=\sigma(j)})_{1 \leq i,j \leq N}.$$

Écrivons les valeurs propres de  $M_{\sigma,z}$

$$\lambda_1(\sigma) = e^{2i\pi\varphi_1(\sigma)}, \dots, \lambda_N(\sigma) = e^{2i\pi\varphi_N(\sigma)},$$

où  $\varphi_1(\sigma), \dots, \varphi_N(\sigma) \in [0, 1[$ . Pour toute fonction 1-périodique  $f : \mathbb{R} \rightarrow \mathbb{R}$  on définit

$$I_{\sigma,N,z}(f) := \sum_{i=1}^N f(\varphi_i(\sigma)).$$

À chaque  $j$ -cycle de  $\sigma$  correspond les valeurs propres de  $M_{\sigma,z}$  de la forme  $T_j \omega$  où  $\omega$  est une racine  $j$ -ème de l'unité et où  $T_j = e^{2i\pi\psi_j}$  est égal en loi à une racine  $j$ -ème de  $\prod_{k=1}^j z_k$ . La permutation  $\sigma$  possédant  $a_{N,j}$  cycles de taille  $j$  pour chaque  $j$ , on a  $a_{N,j}$  variables aléatoires

$$T_j^{(1)}, \dots, T_j^{(a_{N,j})}.$$

On en déduit l'égalité en loi

$$I_{\sigma,N,z}(f) = \sum_{j=1}^N \sum_{l=1}^{a_{N,j}} \sum_{k=0}^{j-1} \tau_{j,l} f\left(\frac{k}{j}\right)$$

où  $\tau_{j,l} f(x) = f\left(x + \psi_j^{(l)}\right)$ . Enfin, pour tout  $j$  on introduit

$$R_{j,l}(f) := \frac{1}{j} \sum_{k=0}^{j-1} \tau_{j,l} f\left(\frac{k}{j}\right) - \int_0^1 \tau_{j,l} f(x) dx = \frac{1}{j} \sum_{k=0}^{j-1} \tau_{j,l} f\left(\frac{k}{j}\right) - \int_0^1 f(x) dx.$$

**Théorème 1.3.2** (Dang, 2012 [33]). *Soit  $\theta > 0$ . Supposons*

$$\sum_{j=1}^{+\infty} j \mathbb{E}(R_{j,1}(f)^2) \in ]0, +\infty[.$$

*Alors sous  $\text{Ewens}(\theta)$ ,  $I_{\sigma,N,z}(f) - \mathbb{E}(I_{\sigma,N,z}(f))$  converge en distribution vers une loi infinidivisible  $\mu$  dont la log-transformée de Fourier est donnée par*

$$\log \hat{\mu}(t) = \sum_{j=1}^{+\infty} \frac{\theta}{j} (\hat{\nu}_j(t) - 1 - itm_j)$$

*où  $\nu_j$  est la loi de  $jR_{j,1}(f)$  et  $m_j$  son espérance.*

**Théorème 1.3.3** (Dang, 2012 [33]). *Soit  $\theta > 0$ . Si  $f$  est une fonction à variation bornée telle que*

$$\sum_{j=1}^{+\infty} j \mathbb{E}(R_{j,1}(f)^2) = +\infty.$$

*Alors sous  $\text{Ewens}(\theta)$ , on a la convergence en distribution*

$$\frac{I_{\sigma,N,z}(f) - \mathbb{E}(I_{\sigma,N,z}(f))}{\sqrt{\text{Var}(I_{\sigma,N,z}(f))}} \longrightarrow \mathcal{N}(0, 1).$$

*De plus,*

$$\mathbb{E}(I_{\sigma,N,z}(f)) = N \int_0^1 f(x) dx + \theta \sum_{j=1}^N \mathbb{E}(R_{j,1}(f)) + \mathcal{O}(1)$$

*et*

$$\text{Var}(I_{\sigma,N,z}(f)) \sim \theta \sum_{j=1}^N j \mathbb{E}(R_j(f)^2).$$

### 1.3.4 Polynômes caractéristiques et fonctions centrales

En 2013, Dang et Zeindler [34] étendent leur résultat au produit en couronne  $S^1 \wr \mathfrak{S}_n$ . Nous présentons leur résultat de manière simplifiée et réduit au cas qui nous intéresse, c'est-à-dire réduit au cas où les entrées non nulles des matrices sont indépendantes et uniformément distribuées sur  $S^1$ .

**Définition 1.3.4.** Soit  $(z_{j,k})_{j,k \geq 1}$  un tableau de variables aléatoires indépendantes uniformément distribuées sur  $S^1$ . Pour toute fonction régulière  $f : S^1 \rightarrow \mathbb{C}$  et  $x \in S^1$  on définit la fonction centrale multiplicative modifiée associée à  $f$

$$\widetilde{W}^{(n)}(f)(x) = \prod_{j=1}^n \prod_{k=1}^{a_{n,j}} f(x^j z_{j,k})$$

et on pose

$$\widetilde{w}^{(n)}(f)(x) = \log(\widetilde{W}^{(n)}(f)(x)) = \sum_{j=1}^n \sum_{k=1}^{a_{n,j}} \log(f(x^j z_{j,k})).$$

**Théorème 1.3.5** (Dang, Zeindler, 2013 [34]). *Soit  $\theta > 0$ . Soit  $f : S^1 \rightarrow \mathbb{C}$  une fonction régulière non nulle et  $x \in S^1$  qui n'est pas une racine de l'unité. Alors, sous  $Ewens(\theta)$ , on a les convergences en distribution*

$$\begin{aligned} \frac{\Re(\widetilde{w}^{(n)}(f)(x))}{\sqrt{\ln n}} - \theta m_R(f) \sqrt{\ln n} &\longrightarrow \mathcal{N}(0, \theta V_R(f)), \\ \frac{\Im(\widetilde{w}^{(n)}(f)(x))}{\sqrt{\ln n}} - \theta m_I(f) \sqrt{\ln n} &\longrightarrow \mathcal{N}(0, \theta V_I(f)) \end{aligned}$$

où

$$\begin{aligned} m_R(f) &= \Re \left( \int_0^1 \log f(e^{2i\pi s}) ds \right), \quad V_R(f) = \int_0^1 \ln^2 |f(e^{2i\pi s})| ds, \\ m_I(f) &= \Im \left( \int_0^1 \log f(e^{2i\pi s}) ds \right), \quad V_I(f) = \int_0^1 \arg^2(f(e^{2i\pi s})) ds. \end{aligned}$$

En particulier en prenant pour  $f$  le polynôme caractéristique  $\widetilde{Z}_n(x) := \det(I - x^{-1} \widetilde{M}_n)$  on obtient :

**Corollaire 1.3.6** (Dang, Zeindler, 2013 [34]). *Soit  $x \in S^1$  qui n'est pas une racine de l'unité. Alors, sous  $Ewens(\theta)$ , les parties réelles et imaginaires de*

$$\frac{\log(\widetilde{Z}_n(x))}{\sqrt{\frac{\pi^2}{12} \theta \ln n}}$$

*convergent chacune en loi vers une variable gaussienne réelle centrée réduite.*

## 1.4 Présentation des résultats

Cette thèse porte sur l'étude des propriétés asymptotiques du spectre des matrices de permutation et des matrices de permutation modifiées. Nous obtenons des résultats pour différentes échelles d'observation du spectre :

- L'échelle *macroscopique* : La taille de la fenêtre d'étude est fixée (ne varie pas avec  $n$ ). Le nombre moyen de valeurs propres observées tend vers l'infini.
- L'échelle *mésoscopique* : On zoome à vitesse  $n\delta_n$  où  $\delta_n$  est une suite de réels strictement positifs tendant vers 0 moins vite que  $1/n$  (de sorte que  $n\delta_n$  tende vers l'infini). Le nombre moyen de valeurs propres observées tend encore vers l'infini.
- L'échelle *microscopique* : On zoome à vitesse  $n$ . Le nombre moyen de valeurs propres observées est borné.

Nos premiers travaux s'inscrivent dans la continuité de l'étude faite par Wieand sur la fonction de comptage des valeurs propres de matrices de permutation suivant la loi uniforme, ainsi que pour les matrices de permutation modifiées (voir Théorèmes 1.2.1 et 1.3.1 ci-avant). Wieand considère des arcs fixes du cercle unité et fait tendre la taille des matrices vers l'infini. Elle obtient que les fluctuations des nombres de valeurs propres dans ces arcs sont asymptotiquement gaussiennes avec une structure de covariance assez particulière. Plusieurs questions naturelles découlent de ce résultat : les fluctuations sont-elles encore asymptotiquement gaussiennes si on s'autorise à faire varier les arcs avec la taille des matrices ? D'autre part, pouvons-nous étendre facilement l'étude à des lois plus générales sur  $\mathfrak{S}_n$  ? Nous montrons que pour tout arc mésoscopique du cercle unité le comportement du nombre de valeurs propres est très analogue à celui pour un arc macroscopique. Nous retrouvons en effet un résultat de normalité asymptotique dans ce cas. De plus, nous généralisons l'étude des arcs macroscopiques et mésoscopiques à la famille de lois d'Ewens. Notamment, pour un nombre fini d'arcs fixés nous dégageons la structure de covariance du vecteur gaussien limite obtenu. La matrice de covariance obtenue ne dépend pas du paramètre  $\theta$  de la loi d'Ewens considérée : c'est exactement la même matrice de covariance que pour la loi uniforme étudiée par Wieand.

Dans un second temps nous nous intéressons au spectre à échelle microscopique. Najnudel et Nikeghbali [68] considèrent la mesure empirique des angles propres normalisés pour des matrices de permutation modifiées ou non et montrent qu'elle converge en un certain sens vers une mesure empirique limite lorsque la taille des matrices tend vers l'infini. Il s'agit en fait d'une convergence de *processus ponctuels* vers un autre processus ponctuel. Partant de ce processus ponctuel limite, nous montrons que les fluctuations du nombre de points dans un intervalle borné  $J$  de  $\mathbb{R}$  sont asymptotiquement gaussiennes lorsque la longueur de  $J$  tend vers l'infini, avec une condition particulière sur la forme de  $J$  pour le cas des matrices de permutation non modifiées. Nous mettons en avant un lien particulier qu'entretiennent les deux processus ponctuels limites associés aux matrices de permutation et aux matrices de permutation modifiées. Enfin, nous étudions le polynôme caractéristique à échelle microscopique en s'inspirant des travaux réalisés par Chhaibi, Najnudel et Nikeghbali [30] sur le CUE. Une fois normalisé de manière appropriée, nous montrons que pour une grande famille de lois sur l'espace des

permutations virtuelles le polynôme caractéristique converge en un certain sens vers une fonction entière aléatoire que nous explicitons. Cette famille de lois englobent les lois d'Ewens.

Nous énonçons maintenant plus précisément nos résultats principaux et esquissons les outils que nous utilisons pour les prouver.

#### 1.4.1 Échelles macroscopique et mésoscopique

Prenons un nombre fini d'arcs du cercle unité  $I_1, \dots, I_m$  de la forme  $I_k := ]e^{2i\pi\alpha_k}, e^{2i\pi\beta_k}]$ . Soit  $\theta > 0$ . Notre premier résultat est le suivant.

**Théorème 1.4.1.** *Sous  $Ewens(\theta)$ , les nombres normalisés de valeurs propres appartenant à  $I_k$*

$$(Y_N^{I_1}, \dots, Y_N^{I_m}) := \left( \frac{X_N^{I_k} - \mathbb{E}(X_N^{I_k})}{\sqrt{\text{Var}(X_N^{I_k})}} \right)_{1 \leq k \leq m}$$

et

$$(\tilde{Y}_N^{I_1}, \dots, \tilde{Y}_N^{I_m}) := \left( \frac{\tilde{X}_N^{I_k} - \mathbb{E}(\tilde{X}_N^{I_k})}{\sqrt{\text{Var}(\tilde{X}_N^{I_k})}} \right)_{1 \leq k \leq m}$$

convergent respectivement en loi vers un vecteur gaussien centré de matrices de covariances respectives  $D = (D_{k,l})_{1 \leq k, l \leq m}$  et  $\tilde{D} = (\tilde{D}_{k,l})_{1 \leq k, l \leq m}$  lorsque  $N$  tend vers l'infini. Ces matrices ne dépendent pas de  $\theta$ , et leurs entrées sont données par

$$D_{k,l} = \frac{c_{k,l}}{\sqrt{c_{k,k}c_{l,l}}}, \quad \tilde{D}_{k,l} = \frac{\tilde{c}_{k,l}}{\sqrt{\tilde{c}_{k,k}\tilde{c}_{l,l}}}$$

où

$$c_{k,l} := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N (\{j\alpha_k\} - \{j\beta_k\})(\{j\alpha_l\} - \{j\beta_l\})$$

et

$$\tilde{c}_{k,l} := \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{j=1}^N (h_j(\beta_k - \alpha_l) + h_j(\alpha_k - \beta_l) - h_j(\alpha_k - \alpha_l) - h_j(\beta_k - \beta_l))$$

avec  $h_j(x) := \{jx\}(1 - \{jx\})$ .

De plus, quand  $N$  tend vers l'infini

$$\text{Var}(X_N^{I_k}) \sim c_{k,k}\theta \ln N, \quad \text{Var}(\tilde{X}_N^{I_k}) \sim \tilde{c}_{k,k}\theta \ln N.$$

En fait ce résultat pour les matrices de permutation sans modification peut être déduit assez directement, sous des conditions supplémentaires légères sur les  $\alpha_k$  et les  $\beta_k$ , du résultat de Dang et Zeindler que nous avons mentionné en introduction (le Corollaire 1.2.10). En effet, écrivons  $e^{i\theta_1}, \dots, e^{i\theta_N}$  les valeurs propres d'une matrice de permutation aléatoire  $M_N$ , où les nombres  $\theta_1, \dots, \theta_N$  ont valeurs dans  $]-\pi, \pi]$ . Soit  $I := ]e^{is}, e^{it}]$

l'arc du cercle unité où  $-\pi < s < t \leq \pi$ . En remarquant que pour tout  $\psi \in ]-\pi, \pi]$ ,

$$\mathbb{1}_{\psi \in ]s, t]} = \frac{t-s}{2\pi} + \frac{1}{\pi} \Im \log(1 - e^{i(\psi-t)}) - \frac{1}{\pi} \Im \log(1 - e^{i(\psi-s)}),$$

alors comme les  $\theta_k$  sont des multiples rationnels de  $2\pi$ , si  $\frac{t}{2\pi}$  est irrationnel on a

$$X_N^I = \sum_{k=1}^N \mathbb{1}_{\theta_k \in ]s, t]} = N \frac{t-s}{2\pi} + \frac{1}{\pi} [\Im \log Z_N(e^{it}) - \Im \log Z_N(e^{is})].$$

Or, d'après le Corollaire 1.2.10, si  $(\frac{s}{2\pi}, \frac{t}{2\pi})$  est de type fini (au sens de (1.1)), on a la convergence en loi

$$\frac{1}{\sqrt{\frac{\pi^2}{12} \theta \ln n}} (\log(Z_N(e^{is})), \log(Z_N(e^{it}))) \longrightarrow (N_1, N_2)$$

où  $\Im(N_1)$  et  $\Im(N_2)$  sont des variables gaussiennes réelles indépendantes centrées réduites. Donc en composant à gauche par l'application continue

$$f : \begin{array}{l} \mathbb{C}^2 \rightarrow \mathbb{R} \\ (z_1, z_2) \mapsto \Im(z_2) - \Im(z_1) \end{array}$$

on obtient une convergence en loi vers une  $\mathcal{N}(0, 2)$ , d'où

$$\frac{X_N^I - \frac{N(t-s)}{2\pi}}{\sqrt{\frac{\theta}{6} \ln n}} \longrightarrow \mathcal{N}(0, 1).$$

Nous retrouvons bien notre résultat de normalité asymptotique pour un arc (pour plusieurs arcs cela fonctionne aussi sous certaines conditions sur les extrémités). En posant  $(\alpha, \beta) = (\frac{s}{2\pi}, \frac{t}{2\pi})$ , la condition  $(\alpha, \beta)$  de type fini implique que  $\alpha$  et  $\beta$  sont des irrationnels n'ayant pas de relation  $\mathbb{Q}$ -linéaire entre eux, et que dans ce cas le calcul de la variance lorsque  $N$  tend vers l'infini donne

$$\text{Var}(X_N^I) \sim \frac{\theta}{6} \log N.$$

À notre connaissance, le spectre des matrices de permutation a été étudié à des échelles macroscopiques et microscopiques, mais jamais à des échelles intermédiaires. Naturellement nous suivons alors la même approche que Wieand mais en considérant un arc non fixé du cercle unité, c'est-à-dire un arc autorisé à varier avec la taille de la matrice. Plus précisément, nous considérons  $I$  de la forme  $I = I_N := (e^{2i\pi\alpha_N}, e^{2i\pi\beta_N}]$ , et faisons l'hypothèse suivante sur  $\delta_N := \beta_N - \alpha_N > 0$  :

$$\begin{cases} \delta_N \xrightarrow{N \rightarrow \infty} 0 \\ N\delta_N \xrightarrow{N \rightarrow \infty} +\infty, \end{cases}$$

ce qui revient à dire que  $I$  est un arc mésoscopique.

**Théorème 1.4.2.** *Sous  $Ewens(\theta)$ , lorsque  $N$  tend vers l'infini on a la convergence en loi*

$$\frac{\tilde{X}_N^I - \mathbb{E}(\tilde{X}_N^I)}{\sqrt{\text{Var}(\tilde{X}_N^I)}} \longrightarrow \mathcal{N}(0, 1)$$

avec

$$\text{Var}(\tilde{X}_N^I) \sim \frac{\theta}{6} \ln(N\delta_N).$$

*En supposant de plus que la suite  $(\alpha_N)$  est constante,  $\alpha_N = \alpha$  pour tout  $N$ , alors lorsque  $N$  tend vers l'infini on a la convergence en loi*

$$\frac{X_N^I - \mathbb{E}(X_N^I)}{\sqrt{\text{Var}(X_N^I)}} \longrightarrow \mathcal{N}(0, 1)$$

avec

$$\text{Var}(X_N^I) \sim \begin{cases} \frac{\theta}{6} \ln(N\delta_N) & \text{si } \alpha \text{ est irrationnel} \\ \theta \left( \frac{1}{6} + \frac{1}{6q^2} \right) \ln(N\delta_N) & \text{si } \alpha = \frac{p}{q} \text{ avec } p, q \text{ premiers entre eux.} \end{cases}$$

Pour prouver ces résultats nous combinons des propriétés du couplage de Feller avec celles de moyennes de Cesàro d'ordres fractionnaires, à la manière de Ben Arous et Dang dans leur étude des fluctuations de statistiques linéaires du spectre. La moyenne de Cesàro d'ordre  $\theta > 0$  d'une suite  $(w_k)_{k \geq 0}$  telle que  $w_0 = 0$  est définie à l'aide des quantités

$$\Psi_n(j) := \frac{n(n-1) \cdots (n-j+1)}{(\theta+n-1) \cdots (\theta+n-j)}, \quad 1 \leq j \leq n,$$

donnant pour la  $n$ -ème moyenne de Cesàro

$$\frac{\theta}{\theta+n} \sum_{j=1}^n \Psi_n(j) w_j.$$

Le cas  $\theta = 1$  est la moyenne de Cesàro au sens classique de la suite  $(w_k)$ . Ce sont les mêmes quantités  $\Psi_{n,j}$  qui apparaissent dans l'expression des premiers moments factoriels des nombres de cycles de mêmes tailles dans une permutation de  $[n]$  suivant une loi d'Ewens de paramètre  $\theta$  : Pour tout  $(r_1, \dots, r_n) \in \mathbb{N}^n$  tel que  $m := r_1 + 2r_2 + \cdots + nr_n \leq n$ , alors

$$\mathbb{E} \left[ \prod_{j=1}^n a_{n,j}^{[r_j]} \right] = \Psi_n(m) \prod_{j=1}^n \left( \frac{\theta}{j} \right)^{r_j}$$

où  $x^{[r]} := x(x-1) \cdots (x-r+1)$  (voir par exemple [4]). Arratia, Barbour et Tavaré font d'ailleurs remarquer que

$$\prod_{j=1}^n \left( \frac{\theta}{j} \right)^{r_j} = \mathbb{E} \left[ \prod_{j=1}^n W_j^{[r_j]} \right]$$

où les  $W_j$  sont des variables de Poisson indépendantes de paramètres  $\frac{\theta}{j}$ .



Le couplage de Feller permet de générer des variables  $a_{n,j}$  et  $W_j$  vérifiant de plus

$$\mathbb{E} \left( \sum_{j=1}^n |a_{n,j} - W_j| \right) = \mathcal{O}_\theta(1) \quad (1.2)$$

Cette majoration est l'ingrédient essentiel pour la démonstration de nos résultats de normalité asymptotique grâce à l'application du théorème de Lindeberg-Feller. Une démonstration de (1.2) est donnée par Arratia Barbour et Tavaré dans [3]. Nous en donnons une démonstration plus directe utilisant des moyennes de Cesàro d'ordres fractionnaires.

D'autre part, pour les calculs du terme principal dans le développement asymptotique de la variance nous utilisons des transformations linéaires permettant notamment d'établir que si la  $n$ -ème moyenne de Cesàro d'ordre  $\theta$  d'une suite de réels positifs  $(w_k)_{k \geq 0}$  converge lorsque  $n$  tend vers l'infini, alors celle-ci est asymptotiquement équivalente à

$$\frac{1}{\ln n} \sum_{j=1}^n \frac{\Psi_n(j)}{j} w_j.$$

Enfin, c'est encore les relations mettant en œuvre les  $\Psi_n(j)$  qui nous permettent d'étudier le comportement asymptotique du plus petit et du plus grand écart entre deux valeurs propres distinctes consécutives. Nous établissons que le plus petit écart a pour ordre de grandeur  $1/n^2$  et le plus grand celui de  $1/n$ . De manière plus précise, si  $d_n$  et  $\mathcal{D}_n$  désignent respectivement le plus petit et plus grand écart entre deux angles propres distincts consécutifs d'une matrice de permutation de taille  $n$  de loi d'Ewens( $\theta$ ), alors nous montrons que les suites  $(n^2 d_n)$ ,  $(1/(n^2 d_n))$ ,  $(n \mathcal{D}_n)$  et  $(1/(n \mathcal{D}_n))$  sont tendues. Autrement dit, on a la proposition suivante :

**Proposition 1.4.3.** *Pour tout  $\varepsilon > 0$ , il existe des constantes  $\nu_\varepsilon, \eta_\varepsilon > 0$  telles que pour tout  $n \geq 1$ ,*

$$\mathbb{P}(\nu_\varepsilon \leq n^2 d_n \leq \eta_\varepsilon) \geq 1 - \varepsilon$$

et

$$\mathbb{P}(\nu_\varepsilon \leq n \mathcal{D}_n \leq \eta_\varepsilon) \geq 1 - \varepsilon.$$

Le même résultat tient pour les matrices de permutation modifiées.

### 1.4.2 Échelle microscopique

Dans le chapitre 3, nous abordons l'étude du comportement asymptotique du spectre d'une manière réciproque à ce que nous avons fait précédemment. Au lieu de regarder les nombres de valeurs propres dans des arcs (fixés ou non) du cercle unité puis de faire tendre la taille de la matrice vers l'infini, nous adoptons une démarche inverse : nous faisons tendre la taille de la matrice vers l'infini avec une normalisation appropriée des angles propres de façon à obtenir un processus ponctuel limite dont l'espacement moyen entre chaque point vaut 1, puis nous regardons le nombre de points dans un intervalle en faisant croître la longueur de cet intervalle jusqu'à l'infini. Le sens d'une telle convergence de processus ponctuels des angles propres normalisés a été rendu rigoureux

par Najnudel et Nikeghbali [68] dans un cadre plus général de modifications de matrices de permutation.

Bien entendu, la seule bonne échelle pour observer un processus ponctuel non trivial à la limite est l'échelle microscopique. En effet, prenons une matrice de permutation aléatoire  $M_n$  (ou modifiée  $\widetilde{M}_n$ ) de taille  $n$ . Elle a  $n$  angles propres dans l'intervalle  $[0, 2\pi[$ , donc l'espacement moyen entre chacun des angles propres vaut  $\frac{2\pi}{n}$ . Ainsi pour maintenir un espacement moyen constant lorsque  $n$  augmente, on doit par exemple multiplier (puis éventuellement translater) les angles propres par  $\frac{1}{2\pi} \times n$  pour avoir un espacement moyen égal à 1.

Soit  $(\sigma_n)_{n \geq 1}$  une suite de permutations suivant la loi d'Ewens( $\theta$ ),  $\theta > 0$ . Pour tout  $n$  nous notons  $(\ell_{n,j})_{j \geq 1}$  la suite des longueurs de cycles de  $\sigma_n$  dans son écriture en cycles à supports disjoints par ordre du plus petit élément qu'ils contiennent, complétée par des zéros, et

$$y_j^{(n)} := \frac{\ell_{n,j}}{n}.$$

Si  $M_n$  et  $\widetilde{M}_n$  correspondent à la permutation  $\sigma_n$ , alors les mesures empiriques de leurs angles propres normalisés par  $\frac{n}{2\pi}$  sont données respectivement par

$$\tau_n = \sum_{j=1}^{+\infty} \mathbb{1}_{y_j^{(n)} > 0} \sum_{k \in \mathbb{Z}} \delta_{\frac{k}{y_j^{(n)}}}$$

et

$$\widetilde{\tau}_n := \sum_{j=1}^{+\infty} \mathbb{1}_{y_j^{(n)} > 0} \sum_{k \in \mathbb{Z}} \delta_{\frac{k + \Phi_{n,j}}{y_j^{(n)}}}$$

où  $(\Phi_{n,j})_{j,n \geq 1}$  est un tableau de variables aléatoires indépendantes et uniformément distribuées sur  $[0, 1]$ . Un résultat de Najnudel et Nikeghbali est que, appliquées à des fonctions continues strictement positives à support compact, ces mesures convergent respectivement vers

$$\tau_\infty := \sum_{j=1}^{+\infty} \sum_{k \in \mathbb{Z} \setminus \{0\}} \delta_{\frac{k}{y_j}}$$

et

$$\widetilde{\tau}_\infty := \sum_{j=1}^{+\infty} \sum_{k \in \mathbb{Z}} \delta_{\frac{k + \Phi_j}{y_j}},$$

où le vecteur aléatoire  $(y_1, y_2, \dots)$  suit la loi GEM de paramètre  $\theta$ , et où les  $\Phi_j$  sont les variables indépendantes de loi uniformes sur  $[0, 1]$ . La convergence est en loi, mais pour les matrices de permutation non modifiées on a aussi une convergence presque sûre si on se place dans le cadre des permutations virtuelles. En effet, si  $(\sigma_n)_{n \geq 1}$  est une permutation virtuelle aléatoire de loi d'Ewens( $\theta$ ), un argument de martingale (voir Tsilevich [85]) montre que pour tout  $j$ ,

$$y_j^{(n)} \longrightarrow y_j$$

presque sûrement quand  $n$  tend vers l'infini, où  $(y_1, y_2, \dots)$  suit la loi GEM de paramètre

$\theta$ . Les coordonnées d'un vecteur suivant la loi GEM ont une représentation à l'aide de produits de  $\text{Beta}(1, \theta)$  indépendantes (voir [4]).

Nous commençons par montrer que le résultat de Najnudel et Nikeghbali s'applique aussi à des indicatrices d'intervalles, de manière à avoir une convergence naturelle du nombre de valeurs propres dans un arc microscopique vers le nombre de points sur l'intervalle correspondant du processus ponctuel limite associé. Plus précisément, nous montrons la proposition suivante :

**Proposition 1.4.4.** *Avec le couplage des permutations virtuelles, pour tous réels positifs  $\alpha < \beta$ ,*

$$\tau_n([\alpha, \beta]) \longrightarrow \tau_\infty([\alpha, \beta])$$

*presque sûrement, et*

$$\tilde{\tau}_n([\alpha, \beta]) \longrightarrow \tilde{\tau}_\infty([\alpha, \beta])$$

*en loi lorsque  $n$  tend vers l'infini.*

Voici maintenant les résultats que nous obtenons à partir des mesures empiriques limites  $\tau_\infty$  et  $\tilde{\tau}_\infty$ .

**Théorème 1.4.5.** *Lorsque le réel  $A$  tend vers l'infini, on a la convergence en loi*

$$\frac{\tilde{\tau}_\infty([0, A]) - A}{\sqrt{\frac{\theta}{6} \log A}} \longrightarrow \mathcal{N}(0, 1).$$

**Théorème 1.4.6.** *Soient  $a, b$  deux réels tels que  $0 < a < b$ . Lorsque  $b$  tend vers l'infini, l'espérance et la variance de  $\tau_\infty([a, a + b])$  s'écrivent*

$$\mathbb{E}(\tau_\infty([a, a + b])) = b - \frac{\theta}{2} \log b + \mathcal{O}_\theta(1),$$

$$\text{Var}(\tau_\infty([a, a + b])) = \frac{\theta}{3} \log b + \mathcal{O}_\theta(\sqrt{\log b}),$$

*et on a la convergence en loi*

$$\frac{\tau_\infty([a, a + b]) - \mathbb{E}(\tau_\infty([a, a + b]))}{\sqrt{\text{Var}(\tau_\infty([a, a + b]))}} \longrightarrow \mathcal{N}(0, 1).$$

*Lorsque  $a$  tend vers l'infini avec  $b = (\nu - 1)a$ , où  $\nu$  est un réel strictement supérieur à 1 fixé quelconque, l'espérance et la variance de  $\tau_\infty([a, \nu a])$  s'écrivent*

$$\mathbb{E}(\tau_\infty([a, \nu a])) = (\nu - 1)a + \mathcal{O}_\theta(1),$$

$$\text{Var}(\tau_\infty([a, \nu a])) = \begin{cases} \frac{\theta}{6} \left(1 - \frac{1}{rs}\right) \log a + \mathcal{O}_\theta(\sqrt{\log a}) & \text{si } \nu = \frac{r}{s} \text{ avec } \text{pgcd}(r, s) = 1 \\ \frac{\theta}{6} \log a + \mathcal{O}_\theta(\sqrt{\log a}) & \text{si } \nu \text{ est irrationnel} \end{cases}$$

*et on a la convergence en loi*

$$\frac{\tau_\infty([a, \nu a]) - \mathbb{E}(\tau_\infty([a, \nu a]))}{\sqrt{\text{Var}(\tau_\infty([a, \nu a]))}} \longrightarrow \mathcal{N}(0, 1).$$

Ce qui est remarquable est que les coefficients apparaissant dans les expressions de ces résultats sont identiques à ceux trouvés dans notre étude à échelles macroscopique et mésoscopique pour des arcs typiques.

Pour le résultat impliquant  $\tilde{\tau}_\infty$ , le mélange des lois rend l'étude assez commode en faisant apparaître la somme infinie

$$\sum_{j=1}^{+\infty} B(p_j)$$

où les  $B(p_j)$  sont des variables de Bernoulli de paramètres  $\{Ay_j\}$ , indépendantes conditionnellement à la suite  $(y_k)_{k \geq 1}$ . Cependant il n'est pas évident d'en tirer une expression asymptotique de la variance (la dépendance des  $y_j$  rend la tâche délicate), c'est pourquoi nous adoptons la stratégie de travailler directement avec la transformée de Fourier de la loi.

Pour le résultat impliquant  $\tau_\infty$ , le principal outil dont nous nous servons est un processus ponctuel de Poisson  $\mathcal{X}$  d'intensité  $\frac{\theta}{x}dx$  sur  $]0, +\infty[$ . En nous inspirant des travaux d'Arratia [2], nous montrons que ce processus ponctuel de Poisson permet de coupler l'ensemble  $\mathcal{V}$  des  $y_j$  avec un autre ensemble de points  $\mathcal{W}$  proche de  $\mathcal{V}$  en un certain sens et plus facile à exploiter. Nous montrons et utilisons que pour toute fonction mesurable bornée  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$\mathbb{E} \left( \left( \sum_{x \in \mathcal{V}} f(x) - \sum_{x \in \mathcal{W}} f(x) \right)^2 \right) \leq C(\theta) \|f\|_\infty^2.$$

Par ailleurs, l'ensemble  $\mathcal{W}$  est égal en loi à  $\mathcal{X} \cap ]0, 1[$ , donc par un théorème de Campbell, la fonction caractéristique de la variable aléatoire  $T := \sum_{x \in \mathcal{W}} f(x)$  a une expression simple dont on déduit, sous certaines hypothèses d'intégrabilité sur  $f$ ,

$$\mathbb{E}(T) = \theta \int_{\mathbb{R}} \frac{f(t)}{t} dt$$

et

$$\text{Var}(T) = \theta \int_{\mathbb{R}} \frac{f(t)^2}{t} dt.$$

Enfin, nous mettons l'accent sur un lien spécial qu'entretiennent les mesures  $\tau_\infty$  et  $\tilde{\tau}_\infty$ . Nous montrons que lorsqu'on translate indéfiniment le processus ponctuel limite associé aux matrices de permutation non modifiées on obtient localement le comportement du processus ponctuel limite associé aux matrices de permutation modifiées. Dit plus formellement, nous avons la proposition suivante :

**Proposition 1.4.7.** *Pour toute fonction  $f \in \mathcal{C}(\mathbb{R}, \mathbb{C})$  à support compact, on a la convergence en loi lorsque  $A$  tend vers l'infini*

$$\langle \tau_\infty \circ T_A, f \rangle \longrightarrow \langle \tilde{\tau}_\infty, f \rangle,$$

où  $T_A$  désigne l'opérateur de translation  $T_A : x \mapsto x + A$ .

Dans le chapitre 4, nous étudions les polynômes caractéristiques respectifs de matrices de permutation et matrices de permutation modifiées

$$Z_n(z) := \det(zI - M_n), \quad \tilde{Z}_n(z) := \det(zI - \tilde{M}_n).$$

Comme nous l'avons mentionné précédemment, nous sortons ici du cadre des lois d'Ewens en prenant des mesures centrales plus générales sur les groupes symétriques, dont l'idée repose sur l'observation suivante :

Pour tout  $n$ , toute mesure centrale sur  $\mathfrak{S}_n$  est entièrement déterminée par une mesure de probabilité sur l'ensemble

$$\nabla^{(n)} := \left\{ (\ell_1, \dots, \ell_n) \in \mathbb{N}^n : \ell_1 \geq \ell_2 \geq \dots \geq \ell_n, \sum_{j=1}^n \ell_j = n \right\}$$

des partitions de l'entier  $n$ , et réciproquement, de telle sorte qu'on a une correspondance biunivoque. Ce que montre Olshanski [74] est que cette correspondance s'étend entre les mesures centrales sur l'espace  $\mathfrak{S}$  des permutations virtuelles et les mesures de probabilité sur

$$\nabla := \left\{ (x_1, x_2, \dots) \in [0, 1]^\infty : x_1 \geq x_2 \geq \dots, \sum_{j=1}^{+\infty} x_j \leq 1 \right\}. \quad (1.3)$$

Afin de caractériser le sous-ensemble de mesures centrales pour lequel nous établissons nos résultats, nous introduisons la définition suivante :

**Définition 1.4.8.** Soit  $p$  une mesure de probabilité sur  $\nabla$ .

- Nous disons que  $p$  est une **mesure à décroissance exponentielle** s'il existe  $r \in (0, 1)$  et  $\nabla_1 \subset \nabla$  avec  $p(\nabla_1) = 1$ , tel que pour tout  $x = (x_1, x_2, \dots) \in \nabla_1$ ,

$$\exists C > 0, \forall j \geq 1, x_j \leq Cr^j. \quad (1.4)$$

- Nous disons qu'une mesure centrale sur  $\mathfrak{S}$  est une **mesure centrale à décroissance exponentielle** si sa mesure associée sur  $\nabla$  est une mesure à décroissance exponentielle.

Avec la Définition 1.4.8, nous remarquons que les lois d'Ewens sont un cas particulier de mesures centrales sur  $\mathfrak{S}$  à décroissance exponentielle, dont les mesures correspondantes sur  $\nabla$  ont support dans

$$\nabla' := \left\{ (x_1, x_2, \dots) \in [0, 1]^\infty : x_1 \geq x_2 \geq \dots, \sum_{j=1}^{+\infty} x_j = 1 \right\} \subset \nabla. \quad (1.5)$$

La méthode que nous employons pour générer les permutations virtuelles à partir de lois sur  $\nabla$  s'inspire grandement des travaux de Tsilevich [85], ainsi que de Najnudel et

Nikeghbali [68] [69]. Elle consiste à tirer un vecteur aléatoire  $y = (y_1, y_2, \dots)$  d'une loi donnée  $p$  sur  $\nabla$ , puis, conditionnellement à  $y$ , de prendre une suite de variables aléatoires  $(x_k)_{k \geq 1}$  indépendantes et uniformément distribuées sur l'ensemble

$$E_y := \bigsqcup_{j=1}^{\infty} \mathcal{C}_j$$

d'union de cercles disjoints  $\mathcal{C}_j$  ayant périmètre  $y_j$ . On peut alors définir une permutation virtuelle  $\sigma_\infty = (\sigma_n)_{n \geq 1}$  où pour tout  $n$ , pour tout  $k \in [n]$ , l'image de  $k$  par  $\sigma_n$  est l'indice du premier point de  $\{x_1, \dots, x_n\}$  rencontré après  $x_k$  lorsqu'on parcourt son cercle dans le sens horaire (c'est  $k$  lui-même si  $x_k$  est le seul point de  $\{x_1, \dots, x_n\}$  du cercle sur lequel il est).

À la différence de l'étude précédente où les longueurs de cycles  $\ell_{n,j}$  étaient considérées par ordre d'apparition des cycles (c'est-à-dire suivant le plus petit élément qu'ils contiennent), ici les longueurs de cycles sont définies par

$$\ell_{n,j} := \#\{k \in \{1, \dots, n\} : x_k \in \mathcal{C}_j\}.$$

Alors, par la loi des grands nombres, conditionnellement à  $y$ , lorsque  $n$  tend vers l'infini on a la convergence presque sûre

$$y_j^{(n)} := \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{x_k \in \mathcal{C}_j} \longrightarrow y_j.$$

Nous considérons les polynômes caractéristiques normalisés

$$\tilde{\xi}_n(z) = \frac{\tilde{Z}_n(e^{2i\pi z/n})}{\tilde{Z}_n(1)}, \quad \xi_{n,\alpha}(z) = \frac{Z_n\left(e^{2i\pi\left(\frac{z}{n} + \alpha\right)}\right)}{Z_n(e^{2i\pi\alpha})}$$

où  $\alpha$  est un nombre irrationnel entre 0 and 1.

Par un couplage analogue à celui des isométries virtuelles introduit par Najnudel et Nikeghbali, nous montrons que le polynôme caractéristique  $\tilde{\xi}_n$  peut s'exprimer grâce à  $(y_j^{(n)})_{j \geq 1}$  et une suite de variables indépendantes  $(u_j)_{j \geq 1}$  uniformément distribuée sur  $S^1$  comme

$$\tilde{\xi}_n(z) = \prod_{\substack{j \geq 1 \\ \ell_{n,j} > 0}} \frac{e^{2i\pi z y_j^{(n)}} - u_j}{1 - u_j}. \quad (1.6)$$

Quant au polynôme caractéristique  $\xi_{n,\alpha}$ , celui-ci s'écrit

$$\xi_{n,\alpha} = \prod_{\substack{j \geq 1 \\ \ell_{n,j} > 0}} \frac{e^{2i\pi\left(\frac{z}{n} + \alpha\right)\ell_{n,j}} - 1}{e^{2i\pi\alpha\ell_{n,j}} - 1}. \quad (1.7)$$

**Théorème 1.4.9.** *Supposons que  $\sigma_\infty$  est générée par une mesure centrale associée à une mesure  $p$  sur  $\nabla'$  à décroissance exponentielle. Alors :*

- *Presque sûrement,  $\tilde{\xi}_n$  converge uniformément sur tout compact vers une fonction*

entière  $\tilde{\xi}_\infty$  définie par

$$\tilde{\xi}_\infty(z) = \prod_{j=1}^{+\infty} \frac{e^{2i\pi z y_j} - u_j}{1 - u_j}.$$

- Supposons que  $\alpha$  est un irrationnel de type fini. Alors

$$\xi_{n,\alpha} \xrightarrow[n \rightarrow \infty]{} \tilde{\xi}_\infty$$

où la convergence est en loi sur l'espace des fonctions continues de  $\mathbb{C}$  dans  $\mathbb{C}$  muni de la topologie de la convergence uniforme sur les compacts.

Remarquons que pour le premier point du Théorème 1.4.9 nous obtenons grâce au couplage des isométries virtuelles une convergence presque sûre, ce qui n'est pas commun en théorie des matrices aléatoires. Sans ce couplage cela devient la même convergence en loi que pour le deuxième point.

De manière similaire à l'étude faite par Chhaibi, Najnudel et Nikeghabli dans [30], nous analysons la fonction limite  $\tilde{\xi}_\infty$  et trouvons que du point de vue des fonctions holomorphes, elle est d'ordre 1. Plus précisément, nous obtenons le résultat suivant :

**Proposition 1.4.10.** *Pour tout  $\varepsilon > 0$ , il existe une constante aléatoire  $C_\varepsilon > 0$  telle que pour tout  $z \in \mathbb{C}$ ,*

$$|\tilde{\xi}_\infty(z)| \leq C_\varepsilon e^{(2\pi+\varepsilon)|z|}$$

et par ailleurs pour tout  $\varepsilon \in (0, 2\pi)$ , il existe une constante aléatoire  $c_\varepsilon > 0$  telle que pour tout  $x \geq 0$ ,

$$|\tilde{\xi}_\infty(-ix)| \geq c_\varepsilon e^{(2\pi-\varepsilon)x}. \quad (1.8)$$

Remarquons que par rapport au CUE où la borne est en  $\exp(\mathcal{O}(z \log z))$ , la borne ici est en  $\exp(\mathcal{O}(z))$ .

Nous terminons notre étude par la suppression de la restriction à  $\nabla'$  pour le support de  $p$ . Nous obtenons encore des convergences de  $\tilde{\xi}_n$  et  $\xi_{n,\alpha}$ , mais cette fois-ci vers deux fonctions entières différentes :

**Théorème 1.4.11.** *Lorsque  $n$  tend vers l'infini*

$$\tilde{\xi}_n(z) \xRightarrow{} \left( \prod_{j=1}^{+\infty} \frac{e^{2i\pi z y_j} - u_j}{1 - u_j} \right) e^{i\pi z(1-y_0)} \prod_{k \in \mathbb{Z}} \left( 1 - \frac{z}{w_k} \right)$$

où  $\{w_k : k \in \mathbb{Z}\}$  sont les points d'un processus de Poisson d'intensité  $1 - y_0$  sur  $\mathbb{R}$ , où  $y_0 := \sum_{j=1}^{+\infty} y_j$  (si  $y_0 = 1$ , nous prenons pour convention  $\prod_{k \in \mathbb{Z}} \left( 1 - \frac{z}{w_k} \right) = 1$ ). Pour tout irrationnel  $\alpha$  de type fini,

$$\xi_{n,\alpha}(z) \xRightarrow{} \left( \prod_{j=1}^{+\infty} \frac{e^{2i\pi z y_j} - u_j}{1 - u_j} \right) e^{i\pi z(1-y_0) \left( 1 - \frac{i}{\tan(\pi\alpha)} \right)}.$$

Le produit  $\prod_{k \in \mathbb{Z}} \left( 1 - \frac{z}{w_k} \right)$  dans le Théorème 1.4.11 n'est pas absolument convergent,

c'est une simplification d'écriture devant être comprise comme

$$\left(1 - \frac{z}{w_0}\right) \prod_{k \geq 1} \left(1 - \frac{z}{w_k}\right) \left(1 - \frac{z}{w_{-k}}\right),$$

où les points du processus de Poisson  $\{w_k : k \in \mathbb{Z}\}$  sont étiquetés comme

$$\cdots < w_{-2} < w_{-1} < 0 \leq w_0 < w_1 < w_2 < \cdots.$$

Nous montrons que les fonctions limites se comportent de façon similaire au cas précédent. En effet :

**Proposition 1.4.12.** *Pour tout  $\varepsilon > 0$ , il existe une constante aléatoire  $C_\varepsilon > 0$  telle que pour tout  $z \in \mathbb{C}$ ,*

$$|\tilde{\xi}_\infty(z)| \leq e^{C_\varepsilon |z| \log(2+|z|)},$$

et

$$|\xi_{\infty, \alpha}(z)| \leq C_\varepsilon e^{(\varepsilon + 2\pi(y_0 + (1-y_0)t_\alpha)|z|)},$$

où  $t_\alpha = \frac{1}{2 \sin(\pi\alpha)} \in (\frac{1}{2}, +\infty)$ .

Les trois prochains chapitres de cette thèse correspondent chacun à un article. Le chapitre 2 contient des résultats à échelles macroscopique et mésoscopique [7], tandis que les chapitres 3 et 4 concernent l'échelle microscopique.



## 2. Counting function of eigenvalues at macroscopic and mesoscopic scales

This chapter is based on the paper *On the number of eigenvalues of modified permutation matrices in mesoscopic intervals*, Journal of Theoretical Probability, (2017), 1-49.

### Contents

<b>2.1</b>	<b>Introduction</b>	<b>31</b>
2.1.1	Random permutation matrices	31
2.1.2	Notations and main results	33
<b>2.2</b>	<b>Preliminaries</b>	<b>35</b>
2.2.1	Cesàro means of fractional order	35
2.2.2	Feller Coupling	38
<b>2.3</b>	<b>Proof of Theorem 1</b>	<b>40</b>
2.3.1	Mean and variance	40
2.3.2	Limiting normality for a finite number of fixed arcs	44
<b>2.4</b>	<b>Proof of Theorem 2</b>	<b>51</b>
2.4.1	Variance	51
2.4.2	Limiting normality for a single mesoscopic arc	59
<b>2.5</b>	<b>Spacing between eigenvalues</b>	<b>62</b>
2.5.1	Largest spacing between two consecutive distinct eigenvalues	62
2.5.2	Smallest spacing between two consecutive distinct eigenvalues	63

### 2.1 Introduction

#### 2.1.1 Random permutation matrices

The spectrum of random permutation matrices has drawn much attention in the last few decades. On the one hand, working with matrices brings a new approach for understanding the structure of permutation groups. On the other hand, the sets of permutation matrices can be seen as finite subgroups of orthogonal groups or unitary groups, and thus, their studies give the opportunity to see how much of the structure of larger groups emerges through these finite subgroups.

To make it clear, let us recall the few following definition and facts:

A **permutation matrix** is a square matrix that has exactly one entry equal to 1 in each row and each column and 0's elsewhere. Since such a matrix is in particular unitary, its spectrum is included in the unit circle. There is a correspondence between  $\mathfrak{S}_N$  (the set of permutations of order  $N$ ) and the set of permutation matrices of size  $N$ . The spectrum of any permutation matrix is completely determined by the **cycle structure** of its corresponding permutation. In other words, this spectrum is a function of the numbers of cycles of same size when one decomposes the permutation into disjoint cycles. Besides, when a permutation is uniformly chosen at random, the joint distribution of these cycle counts is known (see [4], chapter 1). Wieand took advantage of this to investigate the asymptotic behavior of the counting function for the eigenvalues of permutation matrices lying in some fixed arcs on the unit circle [87], and also for some wreath products involving  $\mathfrak{S}_N$  [89], under the uniform distribution. In addition, the work of Blair-Stahn [21] revealed how difficult it is to compute the limiting expectation of the counting function for a shrinking interval of type  $\left(e^{2i\pi a}, e^{2i\pi(a+\frac{b}{N})}\right]$  on the unit circle, with  $a, b$  fixed real numbers, and in particular for some special  $a$  there is not known explicit expression.

The uniform distribution on  $\mathfrak{S}_N$  has not been the only one studied. Indeed, the use of Ewens measures is quite natural [40] and very convenient to study in some mathematical aspects (see [4] and [27] for a definition and related results). Heuristically, the Ewens measures are one-parameter deformations of the uniform distribution, where the parameter (usually denoted by  $\theta > 0$ ) influences the expected total number of cycles in the decomposition into disjoint cycle of a randomly chosen permutation. Ben Arous and Dang [13] tackled this family of measures over permutation matrices and gave some asymptotic results for linear statistics of their spectrum (not only the counting function).

A classical motivation for the study of Ewens measures can be found in population genetics, where the distribution of the  $n$ -tuple whose the  $i$ -th coordinate counts the number of alleles represented  $i$  times in a random sample of  $n$  gametes (taken from a population under certain conditions), is exactly the distribution obtained considering the integer partition induced by a permutation under a certain Ewens measure on  $\mathfrak{S}_n$ . In this framework, the parameter  $\theta$  of the considered Ewens measure plays the role of a *population mutation rate*. (see Ewens' sampling formula [40]).

Furthermore, wreath products have some applications in group theory (*e.g.* finding the exhaustive list of Sylow groups from a given finite permutation group) and in graph theory (*e.g.* looking at automorphisms on regular rooted trees (see for example [39])). Basically, introduction of randomness in these ensembles aims to have a better understanding of their structure.

Before stating in which way we want to extend the results of Wieand and Blair-Stahn in this chapter, let us mention some other relative work, such as the study of characteristic polynomial of random permutation matrices by Hambly, Keevash O'Connell, Stark [48] or of more generally multiplicative class functions for some wreath products by Zeindler *et al.* [92] [93] [34] [35]. We can also mention the various results of Najnudel and Nikeghbali [68] for the point process of eigenvalues where a meaning of almost sure

convergence of the empirical spectral measure is made precise for some modified random permutation matrices.

Now, we introduce the way we shall continue some of the previous works:

- We look at the counting function of eigenvalues for the ensemble of permutation matrices and the wreath product  $S^1 \wr \mathfrak{S}_N$  (where  $S^1$  is the group of complex numbers of modulus 1) endowed with Ewens measures. Our motivation for studying this particular wreath product is twofold: its spectral distribution is quite more convenient to study and it brings closer the analogy with the Circular Unitary Ensemble. Indeed, in contrast with the ensemble of permutation matrices, the distribution of eigenvalues for  $S^1 \wr \mathfrak{S}_N$  is invariant by rotation.
- We take advantage of some tools introduced in the articles of Wieand [87] and Ben Arous, Dang [13] and develop them in our framework.
- We also investigate the counting function at an intermediate scale (mesoscopic) between macroscopic and microscopic scales, where the observed number of eigenvalues still tends to infinity when  $N$  goes to infinity. The study of its fluctuations is motivated by comparison with an analogous result of Bourgade which is given as a consequence of Theorem 1.4 in [23], for unitary matrices.

### 2.1.2 Notations and main results

Let  $\theta > 0$ . Let  $(\sigma_N)_{N \geq 1}$  be a sequence of random permutations following Ewens measure of parameter  $\theta$ . Formally, it means that for all  $N$ ,  $\sigma_N$  takes values in  $\mathfrak{S}_N$  and

$$\forall \sigma \in \mathfrak{S}_N, \mathbb{P}(\sigma_N = \sigma) = \mathbb{P}_\theta^{(N)}(\sigma) = \frac{\theta^{K(\sigma)}}{\theta(\theta+1) \cdots (\theta+N-1)}$$

where  $K(\sigma)$  denotes the total number of cycles of  $\sigma$  once decomposed into disjoint cycles. Let  $(z_j)_{j \geq 1}$  be a sequence of i.i.d random variables uniformly distributed on the unit circle, independent of  $(\sigma_N)_{N \geq 1}$ . For all  $N \geq 1$ , we define  $M_N$  and  $\widetilde{M}_N$  as the  $N$ -by- $N$  matrices whose entries are given by:

$$\forall 1 \leq i, j \leq N, \quad \begin{cases} (M_N)_{i,j} &:= \mathbb{1}_{i=\sigma_N(j)} \\ (\widetilde{M}_N)_{i,j} &:= z_i \mathbb{1}_{i=\sigma_N(j)}. \end{cases}$$

In all the following we identify the ensemble of permutation matrices of order  $N$  and the symmetric group  $\mathfrak{S}_N$ , and consider the outputs of  $\widetilde{M}_N$  as elements of the wreath product of  $S^1$  and  $\mathfrak{S}_N$ , denoted by  $S^1 \wr \mathfrak{S}_N$ .

The elements of both ensembles are in particular unitary matrices, and thus, their eigenvalues belong to the unit circle.

Then, the question of the asymptotic behavior of the distribution of these eigenvalues arises naturally, in particular if one wants to compare them with some known results on other random matrix ensembles.

To this purpose, let  $I := [e^{2i\pi\alpha}, e^{2i\pi\beta}]$  the interval which denotes the arc on the unit circle from  $e^{2i\pi\alpha}$  (excluded) to  $e^{2i\pi\beta}$  (included), with  $0 \leq \alpha < 1$  and  $\alpha < \beta \leq \alpha + 1$ . We

take it half-open for practical reason. The conditions on  $\alpha$  and  $\beta$  are sufficient to take whatever (half-open) interval of the unit circle. For  $N \geq 1$ , we define  $X_N^I$  and  $\tilde{X}_N^I$  as the respective numbers of eigenvalues of  $M_N$  and  $\tilde{M}_N$  lying in  $I$ .

**Lemma 2.1.1.** *Let  $s, t, u, v \in \mathbb{R}$ . The following limits exist, are finite, and can be explicitly computed.*

$$c(s, t, u, v) := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N (\{js\} - \{jt\})(\{ju\} - \{jv\}).$$

$$\tilde{c}(s, t, u, v) := \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{j=1}^N (h_j(t - u) + h_j(s - v) - h_j(s - u) - h_j(t - v))$$

with  $h_j(x) := \{jx\}(1 - \{jx\})$ .

We refer to Lemma 2.3.8 for a condensed version of this result and a proof, inspired from [87] for the first limit and [89] for the second one.

The following theorem have already been established in [87] and [89], in the particular case  $\theta = 1$ . Also, for  $m = 1$  and for random permutation matrices without modification, the result simply derives from Theorem 1.5 in [13]. Furthermore, the second item can be deduced from Corollary 1.2.10 under certain conditions on the family  $(\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_m)$ .

**Theorem 2.1.2.** *Let  $I_1, \dots, I_m$  be a finite number of fixed arcs of the form  $I_k := [e^{2i\pi\alpha_k}, e^{2i\pi\beta_k}]$ . For  $1 \leq k, l \leq m$ , denote  $c_{k,l} := c(\alpha_k, \beta_k, \alpha_l, \beta_l)$  and  $\tilde{c}_{k,l} := \tilde{c}(\alpha_k, \beta_k, \alpha_l, \beta_l)$ . Then, as  $N \rightarrow \infty$*

1.

$$\text{Var}(X_N^{I_k}) \sim c_{k,k} \theta \log N, \quad \text{Var}(\tilde{X}_N^{I_k}) \sim \tilde{c}_{k,k} \theta \log N.$$

2.

$$\left( \frac{X_N^{I_1} - \mathbb{E}(X_N^{I_1})}{\sqrt{\text{Var}(X_N^{I_1})}}, \dots, \frac{X_N^{I_m} - \mathbb{E}(X_N^{I_m})}{\sqrt{\text{Var}(X_N^{I_m})}} \right) \xrightarrow{d} \mathcal{N}(0, D)$$

$$\text{where } D = (D_{k,l})_{1 \leq k, l \leq m} \text{ with } D_{k,l} = \frac{c_{k,l}}{\sqrt{c_{k,k} c_{l,l}}}.$$

3.

$$\left( \frac{\tilde{X}_N^{I_1} - \mathbb{E}(\tilde{X}_N^{I_1})}{\sqrt{\text{Var}(\tilde{X}_N^{I_1})}}, \dots, \frac{\tilde{X}_N^{I_m} - \mathbb{E}(\tilde{X}_N^{I_m})}{\sqrt{\text{Var}(\tilde{X}_N^{I_m})}} \right) \xrightarrow{d} \mathcal{N}(0, \tilde{D})$$

$$\text{where } \tilde{D} = (\tilde{D}_{k,l})_{1 \leq k, l \leq m} \text{ with } \tilde{D}_{k,l} = \frac{\tilde{c}_{k,l}}{\sqrt{\tilde{c}_{k,k} \tilde{c}_{l,l}}}.$$

Most of the innovative work in this chapter holds in the following main result.

**Theorem 2.1.3.** *Assume  $I$  to be depending on  $N$ , of the form  $I = I_N := [e^{2i\pi\alpha_N}, e^{2i\pi\beta_N}]$ .*

Denote  $\delta_N := \beta_N - \alpha_N > 0$ . Suppose that the sequence  $(\delta_N)$  satisfies

$$\begin{cases} \delta_N \xrightarrow{N \rightarrow \infty} 0 \\ N\delta_N \xrightarrow{N \rightarrow \infty} +\infty. \end{cases}$$

1. Then, as  $N \rightarrow \infty$

$$\text{Var}(\tilde{X}_N^I) \sim \frac{\theta}{6} \log(N\delta_N)$$

and

$$\frac{\tilde{X}_N^I - \mathbb{E}(\tilde{X}_N^I)}{\sqrt{\text{Var}(\tilde{X}_N^I)}} \xrightarrow{d} \mathcal{N}(0, 1).$$

2. Suppose in addition that the sequence  $(\alpha_N)$  is constant, say  $\alpha_N = \alpha$  for all  $N$ . Then, as  $N \rightarrow \infty$

$$\text{Var}(X_N^I) \sim \begin{cases} \frac{\theta}{6} \log(N\delta_N) & \text{if } \alpha \text{ is irrational} \\ \theta \left( \frac{1}{6} + \frac{1}{6q^2} \right) \log(N\delta_N) & \text{if } \alpha = \frac{p}{q} \text{ with } p, q \text{ coprime integers} \end{cases}$$

and

$$\frac{X_N^I - \mathbb{E}(X_N^I)}{\sqrt{\text{Var}(X_N^I)}} \xrightarrow{d} \mathcal{N}(0, 1).$$

The chapter is organized as follows: In Section 2.2 we begin with preliminary results about Cesàro means and Feller Coupling. In Sections 2.3 and 2.4 we prove Theorems 2.1.2 and 2.1.3, investigating the asymptotic behavior of the mean and variance of the considered sequences of random variables. In Section 2.5 we look at the extremal spacings between two consecutive eigenvalues and establish some results of tightness. This last section is independent of the Sections 2.3 and 2.4.

## 2.2 Preliminaries

### 2.2.1 Cesàro means of fractional order

We set up here a few results (highly inspired from [95] Volume 1 chapter 3, and [13]) about Cesàro means of fractional order, that we will use to investigate the asymptotical behavior of the variance.

**Definition 2.2.1.** For all  $1 \leq j \leq n$ ,

$$\Psi_n(j) := \frac{n(n-1) \cdots (n-j+1)}{(\theta+n-1) \cdots (\theta+n-j)}. \quad (2.1)$$

**Definition 2.2.2.** 1. The **Cesàro numbers** of order  $\delta \in \mathbb{R} \setminus \{-1, -2, \dots\}$  are given by

$$A_n^\delta := \binom{n+\delta}{n} = \frac{(n+\delta) \cdots (1+\delta)}{n!}.$$

2. The **Cesàro mean** of order  $\theta > 0$  of the sequence  $w = (w_j)_{j \geq 0}$  is given by

$$\sigma_n^\theta(w) := \sum_{j=0}^n \frac{A_{n-j}^{\theta-1}}{A_n^\theta} w_j.$$

3. A sequence of real numbers  $w = (w_j)_{j \geq 0}$  is said to be **convergent in Cesàro sense** of order  $\theta$  (and will be denoted by  $(C, \theta)$ ) to a limit  $\ell$  iff  $\sigma_n^\theta(w) \xrightarrow{n \rightarrow \infty} \ell$ .

*Remark 2.2.3.* If  $w_0 = 0$ , then the Cesàro mean of the sequence  $(w_j)$  can be reformulated as

$$\sigma_n^\theta(w) = \sum_{j=1}^n \frac{A_{n-j}^{\theta-1}}{A_n^\theta} w_j = \frac{\theta}{\theta + n} \sum_{j=1}^n \Psi_n(j) w_j. \quad (2.2)$$

**Lemma 2.2.4** ([95] Vol I, page 74). *Let  $M = (M_{i,j})_{1 \leq j \leq i}$  a lower-triangular infinite stochastic matrix satisfying*

$$\forall j \geq 1, \lim_n M_{n,j} = 0.$$

*Let  $s = (s_1, \dots, s_n, \dots)^T \in \mathbb{R}^\mathbb{N}$  and let  $t = Ms$ . If  $s_n \rightarrow L$ , then  $t_n \rightarrow L$ .*

The next proposition is a particular case of the Lemma 2.27 from [95] Vol II page 70.

**Proposition 2.2.5.** *If a sequence  $(w_n)$  of real numbers is bounded and converging  $(C, 1)$  to a real number  $\ell$ , then it converges  $(C, \delta)$  to  $\ell$  for all  $\delta > 0$ .*

From all this, we can deduce the following results involving the numbers  $\Psi_{n,j}$ :

**Proposition 2.2.6.** *Let  $(w_j)_{j \geq 1}$  be a sequence of non-negative real numbers. Suppose that*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \Psi_n(j) w_j = L > 0.$$

*Then*

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{j=1}^n \frac{\Psi_n(j)}{j} w_j = L\theta.$$

*Proof.* For all  $n \geq 1$ , let us define  $s_n := \frac{1}{n} \sum_{j=1}^n \Psi_n(j) w_j$ , and  $t_n := \frac{1}{J_n} \sum_{j=1}^n \frac{\Psi_n(j)}{j} w_j$ , where

$J_n := \sum_{j=1}^n \frac{\theta}{\theta + j - 1}$ . We introduce the infinite matrix  $M = (M_{n,j})$  defined for all  $n, j \geq 1$  by

$$M_{n,j} = \frac{1}{J_n} \left( \frac{\theta}{\theta + j} \mathbb{1}_{j < n} + \mathbb{1}_{j=n} \right). \quad (2.3)$$

The conditions of Lemma 2.2.4 are easy to check for the matrix  $M$ .

We show  $t_n = \sum_{j=1}^n M_{n,j} s_j$  by induction:

- $n = 1$ :  $t_1 = \Psi_1(1) w_1 = \frac{w_1}{\theta}$  and  $\sum_{j=1}^1 M_{1,j} s_j = s_1 = \frac{w_1}{\theta}$ .

- $n - 1$  to  $n$ : Suppose that the statement holds for  $n - 1$ . We prove it at the step  $n$ , in other words we want to show

$$\sum_{j=1}^n \frac{\Psi_n(j)}{j} w_j = \sum_{j=1}^{n-1} \frac{\theta}{\theta + j} s_j + s_n \quad (2.4)$$

From (2.1) we have

$$\Psi_n(j) \left[ \frac{1}{j} - \frac{1}{n} \right] = \Psi_{n-1}(j) \left[ \frac{1}{j} - \frac{1}{\theta + n - 1} \right]$$

for  $1 \leq j \leq n - 1$  and then, successively,

$$\begin{aligned} \frac{\Psi_n(j)}{j} &= \frac{\Psi_{n-1}(j)}{j} - \frac{\Psi_{n-1}(j)}{\theta + n - 1} + \frac{\Psi_n(j)}{n} \\ \sum_{j=1}^{n-1} \frac{\Psi_n(j)}{j} w_j &= \sum_{j=1}^{n-1} \frac{\Psi_{n-1}(j)}{j} w_j - \frac{1}{\theta + n - 1} \sum_{j=1}^{n-1} \Psi_{n-1}(j) w_j + \frac{1}{n} \sum_{j=1}^{n-1} \Psi_n(j) w_j \\ &= \sum_{j=1}^{n-1} \frac{\Psi_{n-1}(j)}{j} w_j - \frac{n-1}{\theta + n - 1} s_{n-1} + s_n - \frac{\Psi_n(n)}{n} w_n \end{aligned}$$

and, applying the induction hypothesis (2.4) at step  $n - 1$ , we get

$$\begin{aligned} \sum_{j=1}^{n-1} \frac{\Psi_n(j)}{j} w_j + \frac{\Psi_n(n)}{n} w_n &= \sum_{j=1}^{n-2} \frac{\theta}{\theta + j} s_j + s_{n-1} - \frac{n-1}{\theta + n - 1} s_{n-1} + s_n \\ &= \sum_{j=1}^{n-1} \frac{\theta}{\theta + j} s_j + s_n \end{aligned}$$

which is (2.4) at step  $n$ .

Thus Lemma 2.2.4 applies and gives  $t_n \xrightarrow{n \rightarrow \infty} L$ .

Finally, it just remains to see that  $J_n / \log n \rightarrow \theta$ , which is clear by comparison with the harmonic series.  $\square$

**Lemma 2.2.7.** *For all  $n \geq 1$ ,*

$$\frac{1}{n} \sum_{j=1}^n \Psi_n(j) = \frac{1}{\theta} \quad (2.5)$$

and

$$\sum_{j=1}^n \frac{\Psi_n(j)}{j} = \sum_{j=1}^n \frac{1}{\theta + j - 1}. \quad (2.6)$$

*Proof.* First, as  $\sum_{k=0}^{+\infty} A_k^\theta x^k = (1-x)^{-\theta-1} = \frac{1}{1-x} (1-x)^{-(\theta-1)-1}$ , then  $A_n^\theta = \sum_{j=0}^n A_j^{\theta-1}$ . Moreover, it is easy to notice that  $\Psi_n(j) = \frac{\theta+n}{\theta} \frac{A_{n-j}^{\theta-1}}{A_n^\theta}$ ,  $1 \leq j \leq n$ , and  $\frac{\theta+n}{\theta} \frac{A_n^{\theta-1}}{A_n^\theta} = 1$ . Therefore,  $1 + \sum_{j=1}^n \Psi_n(j) = \frac{\theta+n}{\theta}$ , which gives (2.5).

For (2.6), we use the same notation as in the proof of Proposition 2.2.6. We had proven that for all  $n$ ,  $t_n = (Ms)_n$ . In particular, if  $(w_j)$  is the constant sequence equal to 1, then, following (2.5),  $s_j = \frac{1}{\theta}$  for all  $j$ , and

$$\begin{aligned} t_n &= \frac{1}{J_n} \sum_{j=1}^n \frac{\Psi_n(j)}{j} = (Ms)_n = \frac{1}{J_n} \left( \sum_{j=1}^{n-1} \frac{\theta}{\theta+j} s_j + s_n \right) \\ &= \frac{1}{J_n} \left( \sum_{j=1}^{n-1} \frac{1}{\theta+j} + \frac{1}{\theta} \right) = \frac{1}{\theta} \end{aligned}$$

so that we have (2.6). □

### 2.2.2 Feller Coupling

Let  $(\sigma_n)_{n \geq 1}$  be a sequence of random permutation generated under Ewens measure of parameter  $\theta$ . For all  $n \geq 1$  we denote by  $a_{n,j}$  the number of  $j$ -cycles in the decomposition into disjoint cycles of  $\sigma_n$ . The  $a_{n,j}$  are also called the **cycle counts** of  $\sigma_n$ .

The next result will be useful to prove the main results of the chapter. It consists in an approximation of the cycle counts by independent Poisson random variables, using the so-called **Feller Coupling** (see [4]).

**Lemma 2.2.8** (Lemma 5.3 of [4]). *One can couple  $(\sigma_n)_{n \geq 1}$  with a sequence  $(W_j)_{j \geq 1}$  of independent Poisson random variables of parameter  $\theta/j$  in such a way that*

$$\mathbb{E} \left( \sum_{j=1}^n |a_{n,j} - W_j| \right) = \mathcal{O}_\theta(1).$$

*Remark 2.2.9.* A proof of this result is given in [3] pages 525-526, without any consideration of Cesàro numbers. In the Appendix of the present chapter we provide a simple proof involving Cesàro means in order to point out that they naturally emerge from Feller Coupling.

We end preliminaries with two lemmas which will be useful to prove Theorem 2.1.3.

**Lemma 2.2.10.** *For all  $n \geq 1$ ,*

$$\sum_{1 \leq j, k \leq n} \frac{1}{jk} (\Psi_n(j)\Psi_n(k) - \Psi_n(j+k)\mathbb{1}_{j+k \leq n}) = \sum_{k=0}^{n-1} \frac{1}{(\theta+k)^2}. \quad (2.7)$$



*Proof.* Denoting  $K_n := \sum_{j=1}^n a_{n,j}$  the total number of cycles, we first notice that

$$\begin{aligned} \text{Var}(K_n) &= \sum_{j=1}^n \text{Var}(a_{n,j}) + \sum_{\substack{1 \leq j, k \leq n \\ j \neq k}} \text{Cov}(a_{n,j}, a_{n,k}) \\ &= \sum_{1 \leq j, k \leq n} \Psi_n(j+k) \mathbb{1}_{j+k \leq n} \frac{\theta^2}{jk} + \mathbb{E}(K_n) - \sum_{j=1}^n \left( \frac{\theta}{j} \Psi_n(j) \right)^2 \\ &\quad - \sum_{\substack{1 \leq j, k \leq n \\ j \neq k}} \frac{\theta}{j} \Psi_n(j) \frac{\theta}{k} \Psi_n(k) \\ &= \mathbb{E}(K_n) + \theta^2 \sum_{1 \leq j, k \leq n} \frac{1}{jk} (\Psi_n(j+k) \mathbb{1}_{j+k \leq n} - \Psi_n(j) \Psi_n(k)). \end{aligned}$$

Moreover, the Feller Coupling provides the nice expression  $K_n = \xi_1 + \dots + \xi_n$ , where the  $\xi_k$  are independent Bernoulli variables with parameter  $\frac{\theta}{\theta+k-1}$ . From this expression of  $K_n$ , it follows  $\mathbb{E}(K_n) = \sum_{k=0}^{n-1} \frac{\theta}{\theta+k}$  and  $\text{Var}(K_n) = \sum_{k=0}^{n-1} \frac{\theta k}{(\theta+k)^2}$ . Hence,

$$\mathbb{E}(K_n) - \text{Var}(K_n) = \sum_{k=0}^{n-1} \left( \frac{\theta}{\theta+k} - \frac{\theta k}{(\theta+k)^2} \right) = \sum_{k=0}^{n-1} \frac{\theta^2}{(\theta+k)^2},$$

which gives (2.7).  $\square$

**Lemma 2.2.11.** *For all  $n \geq 1$ ,*

$$\sum_{1 \leq j, k \leq n} \frac{1}{jk} |\Psi_n(j) \Psi_n(k) - \Psi_n(j+k) \mathbb{1}_{j+k \leq n}| = \mathcal{O}_\theta(1).$$

*Proof.* Discussing the sign of terms inside absolute values according to  $\theta$ , and using the previous lemma, we get

$$\begin{aligned} &\sum_{1 \leq j, k \leq n} \frac{1}{jk} |\Psi_n(j) \Psi_n(k) - \Psi_n(j+k) \mathbb{1}_{j+k \leq n}| \\ &= \sum_{\substack{1 \leq j, k \leq n \\ j+k \leq n}} \frac{1}{jk} (\Psi_n(j) \Psi_n(k) - \Psi_n(j+k) \mathbb{1}_{j+k \leq n}) (\mathbb{1}_{\theta \geq 1} - \mathbb{1}_{\theta < 1}) \\ &\quad + \sum_{\substack{1 \leq j, k \leq n \\ j+k > n}} \frac{1}{jk} \Psi_n(j) \Psi_n(k) (\mathbb{1}_{\theta \geq 1} + \mathbb{1}_{\theta < 1}) \\ &= (\mathbb{1}_{\theta \geq 1} - \mathbb{1}_{\theta < 1}) \sum_{k=0}^{n-1} \frac{1}{(\theta+k)^2} + 2\mathbb{1}_{\theta < 1} \sum_{\substack{1 \leq j, k \leq n \\ j+k > n}} \frac{1}{jk} \Psi_n(j) \Psi_n(k) \\ &= \mathcal{O}_\theta(1) + 2\mathbb{1}_{\theta < 1} \sum_{\substack{1 \leq j, k \leq n \\ j+k > n}} \frac{1}{jk} \Psi_n(j) \Psi_n(k). \end{aligned}$$

Assume  $\theta < 1$ . It remains to show  $\sum_{\substack{1 \leq j, k \leq n \\ j+k > n}} \frac{1}{jk} \Psi_n(j) \Psi_n(k) = \mathcal{O}_\theta(1)$ .

To do this, we split this sum as the one for  $j$  and  $k$  between  $\frac{n}{10}$  and  $n$ , plus the one for  $j$  or  $k$  between 1 and  $\frac{n}{10}$ . Based on the observation that there exists a constant  $C_\theta$  such that for all  $n$  and  $j \leq n$ ,  $\Psi_n(j) \leq C_\theta \left(\frac{n}{n-j}\right)^{1-\theta}$ , with the convention  $\frac{1}{0^{1-\theta}} = 1$ , it comes

$$\begin{aligned} \sum_{\substack{\frac{n}{10} \leq j, k \leq n \\ j+k > n}} \frac{1}{jk} \Psi_n(j) \Psi_n(k) &\ll_\theta n^{-2\theta} \sum_{\substack{\frac{n}{10} \leq j, k \leq n \\ j+k > n}} \left(\frac{1}{n-j}\right)^{1-\theta} \left(\frac{1}{n-k}\right)^{1-\theta} \\ &= n^{-2\theta} \left( \sum_{0 \leq j \leq \frac{9n}{10}} j^{-(1-\theta)} \right)^2 \ll_\theta 1 \end{aligned}$$

and

$$\begin{aligned} \sum_{\substack{1 \leq j \leq \frac{n}{10} \\ j+k > n}} \frac{1}{jk} \Psi_n(j) \Psi_n(k) &\ll_\theta n^{-\theta} \sum_{1 \leq j \leq \frac{n}{10}} \frac{1}{j} \sum_{k=n-j+1}^n \left(\frac{1}{n-k}\right)^{1-\theta} \\ &\ll_\theta n^{-\theta} \sum_{1 \leq j \leq \frac{n}{10}} \frac{1}{j} j^\theta \ll_\theta 1, \end{aligned}$$

which gives the claim.  $\square$

## 2.3 Proof of Theorem 1

### 2.3.1 Mean and variance

#### Symmetric group $\mathfrak{S}_N$

In order to compare both ensembles with each other, we recall here some known results on the counting function of eigenvalues for  $\mathfrak{S}_N$  that one can find for example in [87] (for the case  $\theta = 1$ ) or in [13]:

**Proposition 2.3.1.** *For all  $N \geq 1$ , denoting  $\omega_j := (\{j\beta\} - \{j\alpha\})$ ,*

$$\mathbb{E}(X_N^I) = N(\beta - \alpha) - \theta \sum_{j=1}^N \frac{\omega_j}{j} \Psi_N(j) \quad (2.8)$$

$$\text{Var}(X_N^I) = \theta \sum_{j=1}^N \frac{\omega_j^2}{j} \Psi_N(j) + \theta^2 \sum_{1 \leq j, k \leq N} \frac{\omega_j \omega_k}{jk} (\Psi_N(j+k) \mathbb{1}_{j+k \leq N} - \Psi_N(j) \Psi_N(k)). \quad (2.9)$$

**Proposition 2.3.2.** *There exists a real number  $c_1 = c_1(\alpha, \beta)$  and a positive real number  $c_2 = c_2(\alpha, \beta)$  such that*

$$\mathbb{E}(X_N^I) \underset{N \rightarrow \infty}{=} N(\beta - \alpha) + c_1 \theta \log N + o(\log N) \quad (2.10)$$

$$\mathrm{Var}(X_N^I) \underset{N \rightarrow \infty}{\sim} c_2 \theta \log N. \quad (2.11)$$

We give two significant examples of values taken by  $c_2$ :

- If  $\alpha$  and  $\beta$  are irrational and linearly independent over  $\mathbb{Q}$ , then  $c_2 = \frac{1}{6}$ .
- If  $\beta$  is irrational and  $\alpha = \frac{p}{q}$  with  $p, q$  coprime numbers, then  $c_2 = \frac{1}{6} + \frac{1}{6q^2}$ .

Details for the computation of the coefficient  $c_2$  and more examples are given in [87]. We complete its study in our Appendix.

### Wreath product $S^1 \wr \mathfrak{S}_N$

To begin with, we give a simple expression of  $\tilde{X}_N^I$  in function of the random variables  $(a_{N,j})_{1 \leq j \leq N}$  and  $(T_{j,p})$ , where the law of  $T_{j,p}$  is the multiplicative convolution of  $j$  independent copies of the uniform distribution on  $S^1$ , *i.e.* the uniform distribution on  $S^1$  itself, and where we recall that  $a_{N,j}$  denotes the number of  $j$ -cycles in  $\sigma_N$ . We have the following equalities in distribution:

$$\tilde{X}_N^I = \sum_{j=1}^N \sum_{p=1}^{a_{N,j}} \sum_{w^j = T_{j,p}} \mathbb{1}_{w \in I} = \sum_{j=1}^N \sum_{p=1}^{a_{N,j}} \sum_{w^j = e^{2i\pi\phi_{j,p}}} \mathbb{1}_{w \in I} \quad (2.12)$$

where the  $(\phi_{j,p})$  are i.i.d random variables, uniformly distributed on  $[0, 1)$ .

Recalling that  $I = (e^{2i\pi\alpha}, e^{2i\pi\beta}]$ , this can be reformulated this way:

$$\begin{aligned} \tilde{X}_N^I &= \sum_{j=1}^N \sum_{p=1}^{a_{N,j}} (j(\beta - \alpha) - \{j\beta - \phi_{j,p}\} + \{j\alpha - \phi_{j,p}\}) \\ &= (\beta - \alpha) \sum_{j=1}^N j a_{N,j} - \sum_{j=1}^N \sum_{p=1}^{a_{N,j}} \left( \{j\beta\} - \{j\alpha\} + \mathbb{1}_{\phi_{j,p} > \{j\beta\}} - \mathbb{1}_{\phi_{j,p} > \{j\alpha\}} \right) \\ &= N(\beta - \alpha) - \sum_{j=1}^N a_{N,j} (\{j\beta\} - \{j\alpha\}) - \sum_{j=1}^N \sum_{p=1}^{a_{N,j}} \left( \mathbb{1}_{\phi_{j,p} > \{j\beta\}} - \mathbb{1}_{\phi_{j,p} > \{j\alpha\}} \right). \end{aligned}$$

*Remark 2.3.3.* It can be noticed that in contrast to the classical ensemble of permutation matrices, if we include the lower endpoint and/or exclude the upper endpoint of the interval, then almost surely the value of the counting function on this interval remains the same.

**Proposition 2.3.4.** *For all  $N \geq 1$ ,*

$$\mathbb{E}(\tilde{X}_N^I) = N(\beta - \alpha) \quad (2.13)$$

$$\mathrm{Var}(\tilde{X}_N^I) = \theta \sum_{j=1}^N \frac{\Psi_N(j)}{j} \{j(\beta - \alpha)\} (1 - \{j(\beta - \alpha)\}). \quad (2.14)$$

*Remark 2.3.5.* For the classical ensemble of permutation matrices, we had a expectation which weakly depended (additional term in  $\log N$ ) on the arithmetic nature of the endpoints of the interval. This is not the case here, and this phenomenon can be well understood since the modification operates uniform random shifts on the sets of eigenangles corresponding to each cycle. Regarding the variance, the effect of endpoints does not vanish since we still have some fractional parts in its expression. More specifically, we have an effect induced by the difference of the endpoints.

*Proof.* First of all we consider the conditional expectation with respect to the random permutation  $\sigma_N$ . We have

$$\mathbb{E}[\tilde{X}_N^I \mid \sigma_N] = N(\beta - \alpha) + \sum_{j=1}^N \sum_{p=1}^N \mathbb{E}[b_{j,p} \mathbf{1}_{a_{N,j} \geq p} \mid \sigma_N]$$

where  $b_{j,p} := \{j\alpha\} - \{j\beta\} + \mathbf{1}_{\phi_{j,p} > \{j\alpha\}} - \mathbf{1}_{\phi_{j,p} > \{j\beta\}}$ .

Since the information of  $\sigma_N$  provides all the information of its cycle structure (*i.e.* the numbers of its cycles of the same sizes), it follows that for all  $j, p$ ,

$$\mathbb{E}[b_{j,p} \mathbf{1}_{a_{N,j} \geq p} \mid \sigma_N] = \mathbf{1}_{a_{N,j} \geq p} \mathbb{E}[b_{j,p} \mid \sigma_N].$$

Moreover the  $b_{j,p}$  are independent of  $\sigma_N$ , hence

$$\mathbb{E}[b_{j,p} \mid \sigma_N] = \mathbb{E}(b_{j,p}) = \{j\alpha\} - \{j\beta\} + \mathbb{P}(\phi_{j,p} > \{j\beta\}) - \mathbb{P}(\phi_{j,p} > \{j\alpha\}) = 0.$$

Consequently all the terms in the double series are zeros. Finally,

$$\mathbb{E}(\tilde{X}_N^I) = \mathbb{E}[\mathbb{E}[\tilde{X}_N^I \mid \sigma_N]] = \mathbb{E}[N(\beta - \alpha)] = N(\beta - \alpha).$$

The computation of the variance is a little longer. Using the fact that the  $b_{j,p}$  are centered and independent,

$$\begin{aligned} \text{Var}(\tilde{X}_N^I) &= \mathbb{E}((\tilde{X}_N^I - \mathbb{E}(\tilde{X}_N^I))^2) \\ &= \mathbb{E} \left( \left[ \sum_{j=1}^N \sum_{p=1}^N \mathbb{E}(b_{j,p} \mathbf{1}_{a_{N,j} \geq p}) \right]^2 \right) \\ &= \sum_{j=1}^N \mathbb{E} \left( \left( \sum_{p=1}^{a_{N,j}} b_{j,p} \right)^2 \right) + 2 \sum_{1 \leq j < k \leq N} \sum_{p=1}^N \sum_{m=1}^N \mathbb{E}(b_{j,p} b_{k,m} \mathbf{1}_{a_{N,j} \geq p} \mathbf{1}_{a_{N,k} \geq m}) \\ &= \sum_{j=1}^N \left[ \sum_{p=1}^N \mathbb{E}(b_{j,p}^2 \mathbf{1}_{a_{N,j} \geq p}) + 2 \sum_{1 \leq m < p \leq N} \mathbb{E}(b_{j,p} b_{j,m} \mathbf{1}_{a_{N,j} \geq p} \mathbf{1}_{a_{N,j} \geq m}) \right] \\ &= \sum_{j=1}^N \sum_{p=1}^N \mathbb{E}(b_{j,p}^2) \mathbb{P}(a_{N,j} \geq p) \end{aligned}$$

with for all  $j, p$ ,

$$\begin{aligned}
 \mathbb{E}(b_{j,p}^2) &= \text{Var}(\mathbb{1}_{\phi_{j,p} > \{j\alpha\}} - \mathbb{1}_{\phi_{j,p} > \{j\beta\}}) \\
 &= \mathbb{E} \left( \mathbb{1}_{\phi_{j,p} > \{j\alpha\}} + \mathbb{1}_{\phi_{j,p} > \{j\beta\}} - 2\mathbb{1}_{\phi_{j,p} > \{j\alpha\}} \mathbb{1}_{\phi_{j,p} > \{j\beta\}} \right) - (\{j\beta\} - \{j\alpha\})^2 \\
 &= 2 - \{j\alpha\} - \{j\beta\} - 2\mathbb{P}(\phi_{j,p} > \max(\{j\alpha\}, \{j\beta\})) - (\{j\beta\} - \{j\alpha\})^2 \\
 &= |\{j\alpha\} - \{j\beta\}| - (\{j\alpha\} - \{j\beta\})^2.
 \end{aligned}$$

Hence,

$$\begin{aligned}
 \text{Var}(\tilde{X}_N^I) &= \sum_{j=1}^N [|\{j\alpha\} - \{j\beta\}| - (\{j\alpha\} - \{j\beta\})^2] \sum_{p=1}^N \mathbb{P}(a_{N,j} \geq p) \\
 &= \sum_{j=1}^N \mathbb{E}(a_{N,j}) [|\{j\alpha\} - \{j\beta\}| - (\{j\alpha\} - \{j\beta\})^2]
 \end{aligned}$$

and we know (see [4] page 96) that for all  $1 \leq j \leq N$ ,

$$\mathbb{E}[a_{N,j}] = \frac{\theta}{j} \cdot \frac{N(N-1) \cdots (N-j+1)}{(\theta+N-1) \cdots (\theta+N-j)} = \frac{\theta}{j} \Psi_N(j). \quad (2.15)$$

It remains to see that for all  $j$ ,  $|\{j\alpha\} - \{j\beta\}| - (\{j\alpha\} - \{j\beta\})^2 = \{j(\beta - \alpha)\}(1 - \{j(\beta - \alpha)\})$ . Indeed, it derives from the next lemma:

**Lemma 2.3.6.** *Let  $x, y$  be real numbers. Then for all  $T \in \mathbb{R}$ ,*

$$|\{x+T\} - \{y+T\}| (1 - |\{x+T\} - \{y+T\}|) = |\{x\} - \{y\}| (1 - |\{x\} - \{y\}|).$$

*Proof.* Without loss of generality we can suppose  $T \in [0, 1)$  (since  $\{x+T\} = \{x + \{T\}\}$ ). We notice first that

$$\{x+T\} - \{y+T\} = \{x\} - \{y\} + \mathbb{1}_{\{y\} > 1-T} - \mathbb{1}_{\{x\} > 1-T}.$$

By discussing the relative positions of  $\{x\}$  and  $\{y\}$  with respect to  $1-T$ , the difference of indicator functions  $\mathbb{1}_{\{y\} > 1-T} - \mathbb{1}_{\{x\} > 1-T}$  takes the value  $-1, 0$  or  $1$ . Thus it is easy to check that in all cases the equality holds.  $\square$

Applying this lemma with  $x = j\beta$  and  $y = -T = j\alpha$  for  $j \in \mathbb{N}^*$ , we deduce (2.14).  $\square$

The next proposition conforms with the intuition one could have as regards with the asymptotic of the variance. We make use of what we set up in preliminaries to prove it.

**Proposition 2.3.7.**

$$\text{Var}(\tilde{X}_N^I) \underset{N \rightarrow \infty}{\sim} \ell \theta \log N \quad (2.16)$$

where

$$\ell := \begin{cases} \frac{1}{6} & \text{if } \beta - \alpha \text{ is irrationnal} \\ \frac{1}{6} - \frac{1}{6q^2} & \text{if } \beta - \alpha = \frac{p}{q} \text{ with } p, q \text{ coprime integers, } q \geq 2. \end{cases}$$

## 2. COUNTING FUNCTION OF EIGENVALUES AT MACROSCOPIC AND MESOSCOPIC SCALES

*Proof.* Let us define  $\delta = \beta - \alpha$  and  $w_j = \{j\delta\}(1 - \{j\delta\})$ ,  $j \in \mathbb{N}^*$ , in such a way that

$$\text{Var}(\tilde{X}_N^I) = \theta \sum_{j=1}^N \frac{\Psi_N(j)}{j} w_j.$$

First, we notice that the sequence  $(w_j)$  is non-negative and bounded (by 1). Moreover, it is proven in [87] that the limits  $\ell_1 := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \{j\delta\}$  and  $\ell_2 := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \{j\delta\}^2$  exist and are finite. Their respective explicit values depend on whether  $\delta$  is rational or irrational. More precisely,

- if  $\delta$  is irrational, then  $\ell_1 = \frac{1}{2}$  and  $\ell_2 = \frac{1}{3}$ ,
- if  $\delta = \frac{p}{q}$  with  $q \geq 1$  and  $\gcd(p, q) = 1$ , then  $\ell_1 = \frac{q(q-1)}{2q^2}$  and  $\ell_2 = \frac{(2q-1)q(q-1)}{6q^3}$ .

Thus, the sequence  $(w_j)$  converges  $(C, 1)$  to  $\ell := \ell_1 - \ell_2 > 0$ . Consequently, we can apply Proposition 2.2.5 on  $w = (w_j)$  so that  $\sigma_N^\theta(w) \xrightarrow{N \rightarrow \infty} \ell$ .

Finally, since  $s_N := \frac{1}{N} \sum_{j=1}^N \Psi_N(j) w_j \underset{N \rightarrow \infty}{\sim} \frac{\sigma_N^\theta(w)}{\theta}$ , then  $s_N \xrightarrow{N \rightarrow \infty} \frac{\ell}{\theta} =: L > 0$ . It follows from Proposition 2.2.6 that

$$\sum_{j=1}^N \frac{\Psi_N(j)}{j} w_j \underset{N \rightarrow \infty}{\sim} L\theta \log N = \ell \log N$$

which implies (2.16). The computation of  $\ell_1$  and  $\ell_2$  is detailed in [87].  $\square$

### 2.3.2 Limiting normality for a finite number of fixed arcs

We consider a finite number of fixed arcs  $I_1, \dots, I_m$  on the unit circle, where  $I_k := (e^{2i\pi\alpha_k}, e^{2i\pi\beta_k}]$ . For  $1 \leq k \leq m$ , we denote by  $c_2^{(k)}$  and  $\ell^{(k)}$  the respective constant numbers appearing in the asymptotic expressions of the variances of  $X_N^{I_k}$  and  $\tilde{X}_N^{I_k}$ . In order to simplify notations, we also define for  $1 \leq k, l \leq m$ ,

$$\omega_{j,k} := \{j\beta_k\} - \{j\alpha_k\}$$

and

$$\begin{aligned} H_{j,k,l} &:= \frac{1}{2} (|\{j\beta_k\} - \{j\alpha_l\}| + |\{j\alpha_k\} - \{j\beta_l\}| \\ &\quad - |\{j\alpha_k\} - \{j\alpha_l\}| - |\{j\beta_k\} - \{j\beta_l\}|) - \omega_{j,k}\omega_{j,l}. \end{aligned}$$

**Lemma 2.3.8.** *The sequences  $(H_{j,k,l})_{j \geq 1}$  and  $(\omega_{j,k}\omega_{j,l})_{j \geq 1}$  converge  $(C, 1)$ .*

*Proof.* One can notice that

$$H_{j,k,l} = \frac{1}{2} (h_j(\alpha_k - \beta_l) + h_j(\beta_k - \alpha_l) - h_j(\alpha_k - \alpha_l) - h_j(\beta_k - \beta_l))$$

where  $h_j(x) := \{jx\}(1 - \{jx\})$ . Moreover, it is clear that for all fixed real numbers  $x$ ,

- if  $x = \frac{p}{q} \in \mathbb{Q}$ , the sequence  $(\{jx\})_{j \geq 1}$  is  $q$ -periodic.

- if  $x \in \mathbb{R} \setminus \mathbb{Q}$ , the sequence  $(\{jx\})_{j \geq 1}$  is equidistributed on  $[0, 1]$ .

Now, if a sequence is periodic then this sequence converges  $(C, 1)$ . Furthermore, if a sequence is equidistributed on  $[0, 1]$  then all continuous functions on  $[0, 1]$  applied to it converge  $(C, 1)$  to the integral of these functions on  $[0, 1]$ . Thus, for all  $x = \alpha_k - \beta_l, \beta_k - \alpha_l, \alpha_k - \alpha_l, \beta_k - \beta_l$ , the sequence  $(h_j(x))_{j \geq 1}$  converges  $(C, 1)$ .

For  $\omega_{j,k}\omega_{j,l}$  it is much more difficult since we have to deal with some products of type  $\{jx\}\{jy\}$  for  $x, y$  real numbers. Discussing the rationality of  $x$  and  $y$  and eventually their linearly dependence over  $\mathbb{Q}$  in the case where they are both irrational, it can be shown that  $(\{jx\}\{jy\})_{j \geq 1}$  converges  $(C, 1)$  to explicit limits, and thus  $(\omega_{j,k}\omega_{j,l})_{j \geq 1}$  converges  $(C, 1)$ . See [87] and our Appendix for details.  $\square$

**Theorem 2.3.9.** Let  $Y_N^{I_k} := \frac{X_N^{I_k} - \mathbb{E}(X_N^{I_k})}{(c_2^{(k)} \theta \log N)^{1/2}}$ , and let  $Z = (Z_1, \dots, Z_m) \sim \mathcal{N}(0, D)$  where  $D$  is the covariance matrix defined for all  $k, l$  by

$$D_{k,l} := \frac{1}{\sqrt{c_2^{(k)} c_2^{(l)}}} \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{j=1}^N \omega_{j,k} \omega_{j,l}.$$

Then

$$(Y_N^{I_1}, \dots, Y_N^{I_m}) \xrightarrow{d} (Z_1, \dots, Z_m).$$

*Proof.* Let  $t := (t_1, \dots, t_m) \in \mathbb{R}^m$ . The theorem will be proven if we show that  $t_1 Y_N^{I_1} + \dots + t_m Y_N^{I_m} \xrightarrow{d} t_1 Z_1 + \dots + t_m Z_m$ . The main idea of the proof is to replace in the expression  $t_1 Y_N^{I_1} + \dots + t_m Y_N^{I_m}$  the  $a_{N,j}$  by independent Poisson random variables  $W_j$  given by the Feller Coupling, show that the difference converges in probability to zero, and then use the new expression to show the convergence in distribution, and finally conclude with Slutsky's theorem. Note that this scheme is very typical when one deals with random permutations, since the approximation of cycle counts by Poisson random variables is natural. Let

$$V_{N,j}^{(t)} := \left( \frac{\theta}{j} - W_j \right) \sum_{k=1}^m \frac{t_k}{(\theta c_2^{(k)} \log N)^{1/2}} \omega_{j,k}.$$

Since the  $W_j$  are independent, then  $(V_{N,j}^{(t)})_{N \geq 1, 1 \leq j \leq N}$  is a triangular array of independent random variables. Let  $T_N^{(t)} := \sum_{j=1}^N V_{N,j}^{(t)}$ . Then

$$\begin{aligned} & t_1 Y_N^{I_1} + \dots + t_m Y_N^{I_m} - T_N^{(t)} \\ &= \sum_{k=1}^m \frac{t_k}{(\theta c_2^{(k)} \log N)^{1/2}} \left[ X_N^{I_k} - \mathbb{E}(X_N^{I_k}) - \sum_{j=1}^N \left( \frac{\theta}{j} - W_j \right) \omega_{j,k} \right] \\ &= \sum_{k=1}^m \frac{t_k}{(\theta c_2^{(k)} \log N)^{1/2}} \sum_{j=1}^N \left[ \frac{\theta}{j} \Psi_N(j) - \frac{\theta}{j} + W_j - a_{N,j} \right] \omega_{j,k}. \end{aligned}$$

We have

$$\begin{aligned} & |t_1 Y_N^{I_1} + \dots + t_m Y_N^{I_m} - T_N^{(t)}| \\ & \leq \sum_{k=1}^m \frac{|t_k|}{(\theta c_2^{(k)} \log N)^{1/2}} \left[ \sum_{j=1}^N \frac{\theta}{j} |\Psi_N(j) - 1| + \sum_{j=1}^N |W_j - a_{N,j}| \right] \end{aligned}$$

with for all  $j$ ,  $|\Psi_N(j) - 1| = (\Psi_N(j) - 1)(\mathbb{1}_{0 < \theta < 1} - \mathbb{1}_{\theta \geq 1})$ , thus using Lemma 2.2.7,

$$\sum_{j=1}^N \frac{\theta}{j} |\Psi_N(j) - 1| = \theta |1 - \theta| \sum_{j=1}^N \frac{1}{j(\theta + j - 1)} = \mathcal{O}_\theta(1).$$

In addition, by Lemma 2.2.8,  $\mathbb{E} \left( \sum_{j=1}^N |W_j - a_{N,j}| \right) = \mathcal{O}_\theta(1)$ .  
Let  $\varepsilon > 0$ . Then, using Markov's inequality,

$$\mathbb{P}(|t_1 Y_N^{I_1} + \dots + t_m Y_N^{I_m} - T_N^{(t)}| > \varepsilon) \leq \frac{1}{\varepsilon (\log N)^{1/2}} \sum_{k=1}^m \frac{|t_k|}{(\theta c_2^{(k)})^{1/2}} \mathcal{O}_\theta(1) \xrightarrow{N \rightarrow \infty} 0.$$

Now, we want to show that  $T_N^{(t)}$  is asymptotically normal. For this purpose, we will check the condition of Lindeberg-Feller on  $(V_{N,j}^{(t)})$ :

$$\lim_{N \rightarrow \infty} \sum_{j=1}^N \mathbb{E} \left( \left( V_{N,j}^{(t)} \right)^2 \mathbb{1}_{|V_{N,j}^{(t)}| > \varepsilon} \right) = 0. \quad (2.17)$$

We have the bound

$$|V_{N,j}^{(t)}| \leq \sum_{k=1}^m \frac{|t_k|}{(\theta c_2^{(k)} \log N)^{1/2}} \left| W_j - \frac{\theta}{j} \right| \leq \frac{C}{(\log N)^{1/2}} \left| W_j - \frac{\theta}{j} \right|$$

where  $C := \sum_{k=1}^m \frac{|t_k|}{(\theta c_2^{(k)})^{1/2}}$ , we deduce that for any  $\varepsilon > 0$ , so that

$$\sum_{j=1}^N \mathbb{E} \left( \left( V_{N,j}^{(t)} \right)^2 \mathbb{1}_{|V_{N,j}^{(t)}| > \varepsilon} \right) \leq \frac{C^2}{\log N} \sum_{j=1}^N \mathbb{E} \left( \left( W_j - \frac{\theta}{j} \right)^2 \mathbb{1}_{\left| W_j - \frac{\theta}{j} \right| > \frac{\varepsilon}{C} (\log N)^{1/2}} \right). \quad (2.18)$$

We could try a fourth moment bound (Lyapunov condition) but it is not sufficient since



$\mathbb{E} \left( \left( W_j - \frac{\theta}{j} \right)^4 \right) = \mathcal{O}(\theta/j)$ . Now, we observe that for all  $j \geq \lceil \theta \rceil$ ,

$$\begin{aligned}
 & \mathbb{E} \left( \left( W_j - \frac{\theta}{j} \right)^2 \mathbb{1}_{\left| W_j - \frac{\theta}{j} \right| > \frac{\varepsilon}{C} (\log N)^{1/2}} \right) \\
 &= \frac{\theta^2}{j^2} \mathbb{1}_{\frac{\theta}{j} > \frac{\varepsilon}{C} (\log N)^{1/2}} e^{-\theta/j} \\
 &+ \left( 1 - \frac{\theta}{j} \right)^2 \mathbb{1}_{\left| 1 - \frac{\theta}{j} \right| > \frac{\varepsilon}{C} (\log N)^{1/2}} \frac{\theta}{j} e^{-\theta/j} \\
 &+ \sum_{k=2}^{+\infty} \left( k - \frac{\theta}{j} \right)^2 \mathbb{1}_{\left| k - \frac{\theta}{j} \right| > \frac{\varepsilon}{C} (\log N)^{1/2}} \frac{\theta^k}{j^k k!} e^{-\theta/j} \\
 &\leq \frac{\theta^2}{j^2} + \frac{\theta}{j} e^{-\theta/j} \mathbb{1}_{1 > \frac{\varepsilon}{C} (\log N)^{1/2}} + \sum_{k=2}^{+\infty} k^2 \frac{\theta^k}{j^k k!} e^{-\theta/j} \\
 &\leq \frac{\theta^2}{j^2} + \mathbb{1}_{1 > \frac{\varepsilon}{C} (\log N)^{1/2}} + \sum_{k=2}^{+\infty} 2k(k-1) \frac{\theta^k}{j^k k!} e^{-\theta/j} \\
 &= \frac{3\theta^2}{j^2} + \mathbb{1}_{1 > \frac{\varepsilon}{C} (\log N)^{1/2}},
 \end{aligned}$$

hence

$$\sum_{j=1}^N \mathbb{E} \left( \left( W_j - \frac{\theta}{j} \right)^2 \mathbb{1}_{\left| W_j - \frac{\theta}{j} \right| > \frac{\varepsilon}{C} (\log N)^{1/2}} \right) \leq N \mathbb{1}_{1 > \frac{\varepsilon}{C} (\log N)^{1/2}} + \mathcal{O}_\theta(1)$$

which, jointly to the bound (2.18) allows to conclude that (2.17) is verified. Consequently the Lindeberg-Feller theorem applies and gives

$$T_N^{(t)} \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

where

$$\begin{aligned}
 \sigma^2 &:= \lim_{N \rightarrow +\infty} \sum_{j=1}^N \mathbb{E} \left( \left( V_{N,j}^{(t)} \right)^2 \right) \\
 &= \lim_{N \rightarrow +\infty} \sum_{j=1}^N \sum_{k,l=1}^m \frac{t_k t_l}{\theta (c_2^{(k)} c_2^{(l)})^{1/2} \log N} \omega_{j,k} \omega_{j,l} \text{Var}(W_j) \\
 &= \sum_{k,l=1}^m \frac{t_k t_l}{(c_2^{(k)} c_2^{(l)})^{1/2}} \lim_{N \rightarrow +\infty} \frac{1}{\log N} \sum_{j=1}^N \frac{\omega_{j,k} \omega_{j,l}}{j}.
 \end{aligned}$$

Finally, since for all  $k, l$ ,  $(\omega_{j,k} \omega_{j,l})_{j \geq 1}$  is bounded and converges  $(C, 1)$  (Lemma 2.3.8), then it follows from Proposition 2.2.6 (taking  $\theta = 1$ ) that

$$\sigma^2 = \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{j=1}^N \sum_{k,l=1}^m \frac{t_k t_l}{(c_2^{(k)} c_2^{(l)})^{1/2}} \omega_{j,k} \omega_{j,l}.$$

Slutsky's theorem ends the proof.  $\square$

**Theorem 2.3.10.** Let  $\tilde{Y}_N^{I_k} := \frac{\tilde{X}_N^{I_k} - \mathbb{E}(\tilde{X}_N^{I_k})}{(\ell^{(k)}\theta \log N)^{1/2}}$ , and let  $\tilde{Z} = (\tilde{Z}_1, \dots, \tilde{Z}_m) \sim \mathcal{N}(0, \tilde{D})$  where  $\tilde{D}$  is the covariance matrix defined for all  $k, l$  by  $\tilde{D}_{k,l} := \frac{1}{\sqrt{\ell^{(k)}\ell^{(l)}}} \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{j=1}^N H_{j,k,l}$ . Then

$$(\tilde{Y}_N^{I_1}, \dots, \tilde{Y}_N^{I_m}) \xrightarrow{d} (\tilde{Z}_1, \dots, \tilde{Z}_m).$$

*Proof.* Let  $t := (t_1, \dots, t_m) \in \mathbb{R}^m$ . Again, it suffices to show that  $t_1 \tilde{Y}_N^{I_1} + \dots + t_m \tilde{Y}_N^{I_m} \xrightarrow{d} t_1 \tilde{Z}_1 + \dots + t_m \tilde{Z}_m$ . In order to shorten the following expressions, let us define for  $j, p \in \mathbb{N}^*$  and  $1 \leq k \leq m$ ,

$$b_{j,p,k} := \mathbb{1}_{\phi_{j,p} > \{j\alpha_k\}} - \mathbb{1}_{\phi_{j,p} > \{j\beta_k\}} - \omega_{j,k}$$

and

$$B_{j,k} := \sum_{p=1}^{W_j} b_{j,p,k}.$$

Let

$$V_{N,j}^{(t)} := \sum_{k=1}^m \frac{t_k}{(\theta \ell^{(k)} \log N)^{1/2}} B_{j,k}.$$

Since the  $W_j$  and the  $\phi_{j,p}$  are independent, then  $(V_{N,j}^{(t)})_{N \geq 1, 1 \leq j \leq N}$  is a triangular array of independent random variables. Let  $T_N^{(t)} := \sum_{j=1}^N V_{N,j}^{(t)}$ . Then

$$\begin{aligned} t_1 \tilde{Y}_N^{I_1} + \dots + t_m \tilde{Y}_N^{I_m} - T_N^{(t)} &= \sum_{k=1}^m \frac{t_k}{(\theta \ell^{(k)} \log N)^{1/2}} \left[ \tilde{X}_N^{I_k} - \mathbb{E}(\tilde{X}_N^{I_k}) - \sum_{j=1}^N B_{j,k} \right] \\ &= \sum_{k=1}^m \frac{t_k}{(\theta \ell^{(k)} \log N)^{1/2}} \sum_{j=1}^N \left[ \sum_{p=1}^{a_{N,j}} b_{j,p,k} - B_{j,k} \right]. \end{aligned}$$

This quantity converges in probability to 0. Indeed, let  $\varepsilon > 0$ . Using Markov's inequality,

$$\begin{aligned} &\mathbb{P}(|t_1 \tilde{Y}_N^{I_1} + \dots + t_m \tilde{Y}_N^{I_m} - T_N^{(t)}| > \varepsilon) \\ &\leq \frac{1}{\varepsilon} \mathbb{E}(|t_1 \tilde{Y}_N^{I_1} + \dots + t_m \tilde{Y}_N^{I_m} - T_N^{(t)}|) \\ &\leq \frac{1}{\sqrt{\log N}} \sum_{k=1}^m \frac{|t_k|}{\varepsilon (\theta \ell^{(k)})^{1/2}} \mathbb{E} \left( \sum_{j=1}^N \left| \sum_{p=1}^{a_{N,j}} b_{j,p,k} - B_{j,k} \right| \right) \end{aligned}$$

with

$$\begin{aligned} \mathbb{E} \left( \sum_{j=1}^N \left| \sum_{p=1}^{a_{N,j}} b_{j,p,k} - B_{j,k} \right| \right) &\leq \mathbb{E} \left( \sum_{j=1}^N \sum_{p=1}^{|a_{N,j}-W_j|} |b_{j,p,k}| \right) \\ &\leq \mathbb{E} \left( \sum_{j=1}^N |a_{N,j} - W_j| \right) \\ &= \mathcal{O}_\theta(1) \end{aligned}$$

applying Lemma 2.2.8. Hence

$$\mathbb{P}(|t_1 \tilde{Y}_N^{I_1} + \dots + t_m \tilde{Y}_N^{I_m} - T_N^{(t)}| > \varepsilon) \xrightarrow{N \rightarrow \infty} 0.$$

Now, we want to show that  $T_N^{(t)}$  is asymptotically normal. For this purpose, we check the condition of Lindeberg-Feller on  $(V_{N,j}^{(t)})$ . Noticing that

$$|V_{N,j}^{(t)}| \leq \sum_{k=1}^m \frac{|t_k|}{(\theta \ell^{(k)} \log N)^{1/2}} |B_{j,k}| \leq \frac{C}{(\log N)^{1/2}} W_j$$

where  $C := \sum_{k=1}^m \frac{|t_k|}{(\theta \ell^{(k)})^{1/2}}$ , we deduce that for any  $\varepsilon > 0$ ,

$$\sum_{j=1}^N \mathbb{E} \left( \left( V_{N,j}^{(t)} \right)^2 \mathbf{1}_{|V_{N,j}^{(t)}| > \varepsilon} \right) \leq \frac{C^2}{\log N} \sum_{j=1}^N \mathbb{E} \left( W_j^2 \mathbf{1}_{W_j > \frac{\varepsilon}{C} (\log N)^{1/2}} \right). \quad (2.19)$$

From this, we observe that for all  $j$ ,

$$\begin{aligned} \mathbb{E} \left( W_j^2 \mathbf{1}_{W_j > \frac{\varepsilon}{C} (\log N)^{1/2}} \right) &= \sum_{k=1}^{+\infty} k^2 \mathbf{1}_{k > \frac{\varepsilon}{C} (\log N)^{1/2}} \mathbb{P}(W_j = k) \\ &= \frac{\theta}{j} e^{-\theta/j} \mathbf{1}_{1 > \frac{\varepsilon}{C} (\log N)^{1/2}} + \mathcal{O}_\theta \left( \frac{1}{j^2} \right) \end{aligned}$$

hence

$$\sum_{j=1}^N \mathbb{E} \left( W_j^2 \mathbf{1}_{W_j > \frac{\varepsilon}{C} (\log N)^{1/2}} \right) \leq \theta \left( \sum_{j=1}^N \frac{1}{j} \right) \mathbf{1}_{1 > \frac{\varepsilon}{C} (\log N)^{1/2}} + \mathcal{O}_\theta(1)$$

which, jointly to the bound (2.19) allows to conclude that (2.17) is verified. Consequently the Lindeberg-Feller theorem applies and gives

$$T_N^{(t)} \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

where

$$\begin{aligned}\sigma^2 &:= \lim_{N \rightarrow +\infty} \sum_{j=1}^N \mathbb{E} \left( \left( V_{N,j}^{(t)} \right)^2 \right) \\ &= \lim_{N \rightarrow +\infty} \sum_{j=1}^N \sum_{k,l=1}^m \frac{t_k t_l}{\theta(\ell^{(k)} \ell^{(l)})^{1/2} \log N} \mathbb{E}(B_{j,k} B_{j,l})\end{aligned}$$

with for all  $j, k, l$ ,

$$\begin{aligned}\mathbb{E}(B_{j,k} B_{j,l}) &= \mathbb{E} \left( \sum_{p,q=1}^{W_j} b_{j,p,k} b_{j,q,l} \right) \\ &= \sum_{p=1}^{+\infty} \mathbb{E}(b_{j,p,k} b_{j,p,l}) \mathbb{P}(W_j \geq p) + 2 \sum_{\substack{p,q=1 \\ p < q}}^{+\infty} \mathbb{E}(b_{j,p,k}) \mathbb{E}(b_{j,q,l}) \mathbb{P}(W_j \geq q) \\ &= \sum_{p=1}^{+\infty} [\mathbb{E}((\mathbb{1}_{\phi_{j,p} > \{j\alpha_k\}} - \mathbb{1}_{\phi_{j,p} > \{j\beta_k\}})(\mathbb{1}_{\phi_{j,p} > \{j\alpha_l\}} - \mathbb{1}_{\phi_{j,p} > \{j\beta_l\}})) \\ &\quad - \omega_{j,k} \omega_{j,l}] \mathbb{P}(W_j \geq p) \\ &= \mathbb{E}(W_j) \left[ \int_0^1 (\mathbb{1}_{x > \{j\alpha_k\}} - \mathbb{1}_{x > \{j\beta_k\}})(\mathbb{1}_{x > \{j\alpha_l\}} - \mathbb{1}_{x > \{j\beta_l\}}) dx - \omega_{j,k} \omega_{j,l} \right] \\ &= \frac{\theta}{j} \left[ \frac{1}{2} (|\{j\beta_k\} - \{j\alpha_l\}| + |\{j\alpha_k\} - \{j\beta_l\}| \right. \\ &\quad \left. - |\{j\alpha_k\} - \{j\alpha_l\}| - |\{j\beta_k\} - \{j\beta_l\}|) - \omega_{j,k} \omega_{j,l} \right] \\ &= \frac{\theta}{j} H_{j,k,l}.\end{aligned}$$

Since for all  $k, l$ , the sequence  $(H_{j,k,l})_{j \geq 1}$  is bounded (by 1) and converges  $(C, 1)$  to a finite limit (Lemma 2.3.8), then it follows from Proposition 2.2.6 (taking  $\theta = 1$ ) that

$$\sigma^2 = \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{j=1}^N \sum_{k,l=1}^m \frac{t_k t_l}{(\ell^{(k)} \ell^{(l)})^{1/2}} H_{j,k,l}.$$

Slutsky's theorem ends the proof.  $\square$

Now, we can give a significant particular case of the two previous theorems:

**Corollary 2.3.11.** *With the same notation, let assume the  $\alpha_k$  and the  $\beta_l$ ,  $1 \leq k, l \leq m$ , to be irrational numbers which are linearly independent over  $\mathbb{Q}$ . Then  $D$  and  $\tilde{D}$  are the identity matrix.*

Indeed, this corollary is a significant version since the additional condition is almost surely satisfied if the endpoints of the intervals are uniformly sampled on the unit circle.

*Proof.* Let  $1 \leq k, l \leq m$ . Under the assumption, it suffices to notice that

- If  $k \neq l$ ,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \omega_{j,k} \omega_{j,l} = \int_0^1 \int_0^1 \int_0^1 \int_0^1 (x-y)(z-t) dx dy dz dt = 0,$$

and

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N H_{j,k,l} = 0$$

thus  $D_{k,l} = 0$  and  $\tilde{D}_{k,l} = 0$ .

- If  $k = l$ ,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N H_{j,k,l} + \omega_{j,k} \omega_{j,l} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N |\omega_{j,k}| = \int_0^1 \int_0^1 |x-y| dx dy = \frac{1}{3}$$

and

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \omega_{j,k}^2 = \int_0^1 \int_0^1 (x-y)^2 dx dy = \frac{1}{6},$$

thus  $D_{k,l} = 6 \times \frac{1}{6} = 1$  and  $\tilde{D}_{k,l} = 6 \left( \frac{1}{3} - \frac{1}{6} \right) = 1$ .

□

## 2.4 Proof of Theorem 2

### 2.4.1 Variance

Here, we assume that the interval  $I$  shrinks as  $N$  tends to infinity. For  $N \geq 1$  we define  $I_N := (e^{2i\pi\alpha_N}, e^{2i\pi\beta_N}]$ , and  $\delta_N := \beta_N - \alpha_N \in (0, 1]$ .

**Proposition 2.4.1.** *Suppose that the sequence  $(\delta_N)$  satisfies* 
$$\begin{cases} \delta_N \xrightarrow{N \rightarrow \infty} 0 \\ N\delta_N \xrightarrow{N \rightarrow \infty} +\infty. \end{cases} \quad \text{Then}$$

$$\text{Var}(\tilde{X}_N^{I_N}) \underset{N \rightarrow \infty}{\sim} \frac{\theta}{6} \log(N\delta_N). \quad (2.20)$$

*Proof.* First, with respect to what we have stated before,

$$\text{Var}(\tilde{X}_N^{I_N}) = \theta \sum_{j=1}^N \frac{\Psi_N(j) \chi_N(j)}{j}$$

where  $\chi_N(j) := \{j\delta_N\} (1 - \{j\delta_N\})$ . We are going to study the particular case  $\theta = 1$ , then we will see that the general case  $\theta > 0$  can be quickly deduced from it.

#### Particular case $\theta = 1$ :

Since the function  $t \mapsto \{t\}$  is of period 1, then  $t \mapsto \chi_N(t)$  is periodic of period  $1/\delta_N$ . Thus, an idea may be to group the terms in the sum according to  $1/\delta_N$ . We begin with

## 2. COUNTING FUNCTION OF EIGENVALUES AT MACROSCOPIC AND MESOSCOPIC SCALES

---

separating the points of the first period from those of the others:

$$\text{Var}(\tilde{X}_N^{I_N, \theta=1}) = \sum_{j=1}^N \frac{\chi_N(j)}{j} = \sum_{j=1}^{\lceil 1/\delta_N \rceil - 1} \frac{\chi_N(j)}{j} + \sum_{j=\lceil 1/\delta_N \rceil}^N \frac{\chi_N(j)}{j}$$

where

$$\begin{aligned} \sum_{j=1}^{\lceil 1/\delta_N \rceil - 1} \frac{\chi_N(j)}{j} &= \sum_{j=1}^{\lceil 1/\delta_N \rceil - 1} \frac{1}{j} j \delta_N (1 - j \delta_N) \\ &= \delta_N \sum_{j=1}^{\lceil 1/\delta_N \rceil - 1} (1 - j \delta_N) \xrightarrow{N \rightarrow \infty} \int_0^1 (1 - x) dx = \frac{1}{2}. \end{aligned} \tag{2.21}$$

We apply a summation by parts on the second sum (Abel transformation):

$$\begin{aligned} \sum_{j=\lceil 1/\delta_N \rceil}^N \frac{\chi_N(j)}{j} &= \frac{1}{N} \sum_{j=\lceil 1/\delta_N \rceil}^N \chi_N(j) - \sum_{j=\lceil 1/\delta_N \rceil}^{N-1} \left( \frac{1}{j+1} - \frac{1}{j} \right) \sum_{k=\lceil 1/\delta_N \rceil}^j \chi_N(k) \\ &= \frac{1}{N} \sum_{j=\lceil 1/\delta_N \rceil}^N \chi_N(j) + \sum_{j=\lceil 1/\delta_N \rceil}^{N-1} \frac{1}{j(j+1)} \sum_{k=\lceil 1/\delta_N \rceil}^j \chi_N(k) \end{aligned}$$

Let  $j \geq \lceil 1/\delta_N \rceil$ , and define

$$A_j := \frac{1}{j} \sum_{p=1}^{\lfloor j\delta_N \rfloor - 1} \sum_{k=\lceil p/\delta_N \rceil}^{\lceil (p+1)/\delta_N \rceil - 1} \chi_N(k), \quad B_j := \frac{1}{j} \sum_{k=\lceil j\delta_N \rceil / \delta_N}^j \chi_N(k)$$

in such a way that  $A_j + B_j = \frac{1}{j} \sum_{k=\lceil 1/\delta_N \rceil}^j \chi_N(k)$ . On the one hand, we observe that

$$|B_j| \leq \frac{1}{j} (j - \lceil j\delta_N \rceil / \delta_N + 1) \leq \frac{1 + 1/\delta_N}{j}$$

so

$$\begin{aligned} \left| \sum_{j=\lceil 1/\delta_N \rceil}^{N-1} \frac{B_j}{j+1} \right| &\leq (1 + 1/\delta_N) \sum_{j=\lceil 1/\delta_N \rceil}^{N-1} \frac{1}{j(j+1)} = (1 + 1/\delta_N) \left( \frac{1}{\lceil 1/\delta_N \rceil} - \frac{1}{N} \right) \\ &= \mathcal{O}(1). \end{aligned}$$

On the other hand,

$$A_j = \frac{1}{j} \sum_{p=1}^{\lfloor j\delta_N \rfloor - 1} \sum_{q=0}^{\lceil (p+1)/\delta_N \rceil - 1 - \lceil p/\delta_N \rceil} \chi_N(q + \lceil p/\delta_N \rceil)$$

where

$$\begin{aligned} \lceil (p+1)/\delta_N \rceil - 1 - \lfloor p/\delta_N \rfloor &= 1/\delta_N + \{-(p+1)/\delta_N\} - \{-p/\delta_N\} - 1 \\ &= \lfloor 1/\delta_N \rfloor - \mathbf{1}_{\{-p/\delta_N\} \geq \{-(p+1)/\delta_N\}} \end{aligned}$$

and

$$\chi_N(q + \lceil p/\delta_N \rceil) = \chi_N(q + \{-p/\delta_N\}).$$

Since for all  $q \in \llbracket 0, \lceil (p+1)/\delta_N \rceil - 1 - \lfloor p/\delta_N \rfloor \rrbracket$ ,

$$0 \leq \delta_N(q + \{-p/\delta_N\}) \leq \delta_N(1/\delta_N + \{-(p+1)/\delta_N\} - 1) < 1$$

then  $\{\delta_N(q + \{-p/\delta_N\})\} = \delta_N(q + \{-p/\delta_N\})$ , and

$$A_j = \frac{1}{j} \sum_{p=1}^{\lfloor j\delta_N \rfloor - 1} \sum_{q=0}^{\lfloor 1/\delta_N \rfloor - \mathbf{1}_{\{-p/\delta_N\} \geq \{-(p+1)/\delta_N\}}} \delta_N(q + \{-p/\delta_N\})(1 - \delta_N(q + \{-p/\delta_N\})).$$

Let us introduce

$$\tilde{A}_j := \frac{1}{j} (\lfloor j\delta_N \rfloor - 1) \sum_{q=1}^{\lfloor 1/\delta_N \rfloor} q\delta_N(1 - q\delta_N).$$

We have for all  $j \geq \lceil 1/\delta_N \rceil$ ,

$$\begin{aligned} |A_j - \tilde{A}_j| &= \left| \frac{1}{j} \sum_{p=1}^{\lfloor j\delta_N \rfloor - 1} \sum_{q=0}^{\lfloor 1/\delta_N \rfloor - \mathbf{1}_{\{-p/\delta_N\} \geq \{-(p+1)/\delta_N\}}} \delta_N\{-p/\delta_N\} (1 - 2q\delta_N - \{-p/\delta_N\}\delta_N) \right. \\ &\quad \left. - \mathbf{1}_{\{-p/\delta_N\} \geq \{-(p+1)/\delta_N\}} \delta_N \lfloor 1/\delta_N \rfloor (1 - \delta_N \lfloor 1/\delta_N \rfloor) \right| \\ &\leq \frac{1}{j} \sum_{p=1}^{\lfloor j\delta_N \rfloor - 1} \sum_{q=0}^{\lfloor 1/\delta_N \rfloor - \mathbf{1}_{\{-p/\delta_N\} \geq \{-(p+1)/\delta_N\}}} \delta_N\{-p/\delta_N\} |1 - 2q\delta_N - \{-p/\delta_N\}\delta_N| \\ &\quad + \mathbf{1}_{\{-p/\delta_N\} \geq \{-(p+1)/\delta_N\}} \delta_N \lfloor 1/\delta_N \rfloor (1 - \delta_N \lfloor 1/\delta_N \rfloor) \\ &\leq \frac{1}{j} (\lfloor j\delta_N \rfloor - 1) [2\delta_N(1 + 1/\delta_N) + \delta_N] \leq 2\delta_N^2(1/\delta_N + 2) \end{aligned}$$

hence

$$\begin{aligned} \left| \sum_{j=\lceil 1/\delta_N \rceil}^{N-1} \frac{A_j}{j+1} - \sum_{j=\lceil 1/\delta_N \rceil}^{N-1} \frac{\tilde{A}_j}{j+1} \right| &\leq 2\delta_N^2(1/\delta_N + 2) \sum_{j=\lceil 1/\delta_N \rceil}^{N-1} \frac{1}{j+1} \\ &\stackrel{N \rightarrow \infty}{\sim} 2\delta_N \log(N\delta_N) = o(\log(N\delta_N)). \end{aligned}$$

Moreover,

$$\sum_{j=\lceil 1/\delta_N \rceil}^{N-1} \frac{\tilde{A}_j}{j+1} = \delta_N \sum_{q=1}^{\lfloor 1/\delta_N \rfloor} q\delta_N(1 - q\delta_N) \sum_{j=\lceil 1/\delta_N \rceil}^{N-1} \left( \frac{1}{j+1} - \frac{(1 + \{j\delta_N\})/\delta_N}{j(j+1)} \right)$$

with

$$\delta_N \sum_{q=1}^{\lfloor 1/\delta_N \rfloor} q\delta_N (1 - q\delta_N) \xrightarrow{N \rightarrow \infty} \int_0^1 x(1-x)dx = \frac{1}{6}$$

and

$$\sum_{j=\lfloor 1/\delta_N \rfloor}^{N-1} \left( \frac{1}{j+1} - \frac{(1 + \{j\delta_N\})/\delta_N}{j(j+1)} \right) \underset{N \rightarrow \infty}{\sim} \log(N\delta_N)$$

since

$$\left| \sum_{j=\lfloor 1/\delta_N \rfloor}^{N-1} \frac{(1 + \{j\delta_N\})/\delta_N}{j(j+1)} \right| \leq \frac{2}{\delta_N} \sum_{j=\lfloor 1/\delta_N \rfloor}^{N-1} \frac{1}{j(j+1)} = \frac{2}{\delta_N} \left( \frac{1}{\lfloor 1/\delta_N \rfloor} - \frac{1}{N} \right) = \mathcal{O}(1).$$

Consequently,

$$\sum_{j=\lfloor 1/\delta_N \rfloor}^{N-1} \frac{A_j}{j+1} \underset{N \rightarrow \infty}{\sim} \frac{1}{6} \log(N\delta_N).$$

Furthermore we notice that  $|\tilde{A}_N| \leq \delta_N \sum_{q=1}^{\lfloor 1/\delta_N \rfloor} q\delta_N (1 - q\delta_N) = \frac{1}{6} + o(1)$ , thus  $A_N + B_N = \mathcal{O}(1)$ . We deduce

$$\sum_{j=\lfloor 1/\delta_N \rfloor}^N \frac{\chi_N(j)}{j} = A_N + B_N + \sum_{j=\lfloor 1/\delta_N \rfloor}^{N-1} \frac{A_j + B_j}{j+1} \underset{N \rightarrow \infty}{\sim} \frac{1}{6} \log(N\delta_N) \quad (2.22)$$

Finally, combining (2.21) and (2.22) we get (2.20) for the case  $\theta = 1$ .

**General case  $\theta > 0$ :**

By triangular inequality, we have

$$\left| \sum_{j=1}^N \frac{\Psi_N(j)\chi_N(j)}{j} - \sum_{j=1}^N \frac{\chi_N(j)}{j} \right| \leq \sum_{j=1}^N \frac{1}{j} |\Psi_N(j) - 1|.$$

Moreover, as for any given  $\theta > 0$  all the terms  $\Psi_N(j) - 1$  have constant sign (the sign of  $1 - \theta$ ), then, using Lemma 2.2.7, it follows

$$\sum_{j=1}^N \frac{1}{j} |\Psi_N(j) - 1| = |\theta - 1| \sum_{j=1}^N \frac{1}{j(\theta + j - 1)}$$

which converges as  $N$  goes to infinity, therefore

$$\sum_{j=1}^N \frac{\Psi_N(j)\chi_N(j)}{j} = \sum_{j=1}^N \frac{\chi_N(j)}{j} + \mathcal{O}_\theta(1).$$



We deduce from (2.21) and (2.22) that

$$\sum_{j=1}^N \frac{\Psi_N(j)\chi_N(j)}{j} \underset{N \rightarrow \infty}{=} \frac{1}{6} \log(N\delta_N) + o(\log N\delta_N),$$

which gives (2.20) for all  $\theta > 0$ .  $\square$

**Proposition 2.4.2.** *Suppose that the sequence  $(\delta_N)$  satisfies  $\begin{cases} \delta_N \xrightarrow{N \rightarrow \infty} 0 \\ N\delta_N \xrightarrow{N \rightarrow \infty} +\infty. \end{cases}$  Suppose in addition that  $(\alpha_N)$  is a constant sequence, say  $\alpha_N = \alpha$  for all  $N$ . Then*

$$\text{Var}(X_N^{I_N}) \underset{N \rightarrow \infty}{\sim} \begin{cases} \frac{\theta}{6} \log(N\delta_N) & \text{if } \alpha \text{ is irrational} \\ \theta \left( \frac{1}{6} + \frac{1}{6q^2} \right) \log(N\delta_N) & \text{if } \alpha = \frac{p}{q} \text{ with } p, q \text{ coprime integers, } q \geq 1. \end{cases} \quad (2.23)$$

*Proof.* For  $1 \leq j \leq N$ , let  $\omega_j^{(N)} := \{j\beta_N\} - \{j\alpha_N\} = \{j(\alpha + \delta_N)\} - \{j\alpha\}$ . We recall that the variance at step  $N$  is given by

$$\begin{aligned} \text{Var}(X_N^{I_N}) &= \theta \sum_{j=1}^N \frac{(\omega_j^{(N)})^2}{j} \Psi_N(j) \\ &\quad + \theta^2 \sum_{1 \leq j, k \leq N} \frac{\omega_j^{(N)} \omega_k^{(N)}}{jk} (\Psi_N(j+k) \mathbb{1}_{j+k \leq N} - \Psi_N(j) \Psi_N(k)). \end{aligned}$$

Since all the  $\omega_j^{(N)}$  are bounded by 1, we have

$$\begin{aligned} &\left| \sum_{1 \leq j, k \leq N} \frac{\omega_j^{(N)} \omega_k^{(N)}}{jk} (\Psi_N(j+k) \mathbb{1}_{j+k \leq N} - \Psi_N(j) \Psi_N(k)) \right| \\ &\leq \sum_{1 \leq j, k \leq N} \frac{1}{jk} |\Psi_N(j+k) \mathbb{1}_{j+k \leq N} - \Psi_N(j) \Psi_N(k)| \\ &= \mathcal{O}_\theta(1) \end{aligned}$$

by Lemma 2.2.11, so that

$$\text{Var}(X_N^{I_N}) = \theta \sum_{j=1}^N \frac{(\omega_j^{(N)})^2}{j} \Psi_N(j) + \mathcal{O}_\theta(1). \quad (2.24)$$

Assume  $\alpha$  to be irrational. The result derives from the following lemma:

**Lemma 2.4.3.** *Let  $t \in \mathbb{R} \setminus \mathbb{Q}$ . Let  $(\varepsilon_n)_{n \geq 1}$  be a sequence of positive real numbers which converges to 0. Let  $f$  be a real-valued continuous function on  $[0, 1]$ . Then*

$$\varepsilon_n \sum_{j=1}^{\lceil \frac{1}{\varepsilon_n} \rceil - 1} f(j\varepsilon_n) \mathbb{1}_{j\varepsilon_n \geq 1 - \{jt\}} \xrightarrow{n \rightarrow \infty} \int_0^1 x f(x) dx.$$

## 2. COUNTING FUNCTION OF EIGENVALUES AT MACROSCOPIC AND MESOSCOPIC SCALES

---

*Proof.* For all  $n \geq 1$ , let  $\mu_n := \varepsilon_n \sum_{j=1}^{\lceil \frac{1}{\varepsilon_n} \rceil - 1} \delta_{(j\varepsilon_n, jt)}$  a measure on the torus  $\mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}$ . For all  $(k, l) \in \mathbb{Z}^2$ , the Fourier transform of  $\mu_n$  in  $(k, l)$  is given by

$$\widehat{\mu}_n(k, l) = \varepsilon_n \sum_{j=1}^{\lceil \frac{1}{\varepsilon_n} \rceil - 1} \left( e^{2i\pi(k\varepsilon_n + lt)} \right)^j$$

which converges to 1 if  $(k, l) = (0, 0)$  and to 0 otherwise as  $n$  goes to infinity. Thus  $(\mu_n)$  converges weakly to the uniform measure on  $\mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}$ . Let  $f$  be a real-valued continuous function on  $[0, 1]$  and let  $g$  be a function from the torus to  $\mathbb{R}$  defined by  $g(x, y) = f(x)\mathbb{1}_{x \geq 1-y}$ . Then  $g$  is continuous everywhere excepted at most on  $x = 0$ ,  $y = 0$  and  $x = 1 - y$ . Consequently, the set of discontinuities of  $g$  is at most one-dimensional, which is of measure zero for the Lebesgue measure of dimension 2. Hence

$$\lim_{n \rightarrow \infty} \varepsilon_n \sum_{j=1}^{\lceil \frac{1}{\varepsilon_n} \rceil - 1} g(j\varepsilon_n, jt) = \int_0^1 \int_0^1 g(x, y) dx dy = \int_0^1 x f(x) dx.$$

□

Noticing that  $\omega_j^{(N)} = \{j\delta_N\} - \mathbb{1}_{\{j\delta_N\} \geq 1 - \{j\alpha\}}$  and following the same scheme than previously (treating the case  $\theta = 1$  and then the case  $\theta > 0$ ), we get

$$\begin{aligned} \theta \sum_{j=1}^N \frac{(\omega_j^{(N)})^2}{j} \Psi_N(j) &\underset{N \rightarrow \infty}{=} \theta \log(N\delta_N) \delta_N \sum_{q=1}^{\lceil \frac{1}{\delta_N} \rceil} (q\delta_N - \mathbb{1}_{q\delta_N \geq 1 - \{q\alpha\}})^2 \\ &\quad + o(\log(N\delta_N)) \end{aligned}$$

with

$$\begin{aligned} \delta_N \sum_{q=1}^{\lceil \frac{1}{\delta_N} \rceil} (q\delta_N - \mathbb{1}_{q\delta_N \geq 1 - \{q\alpha\}})^2 &= \delta_N \sum_{q=1}^{\lceil \frac{1}{\delta_N} \rceil} [(q\delta_N)^2 + (1 - 2q\delta_N)\mathbb{1}_{q\delta_N \geq 1 - \{q\alpha\}}] \\ &\xrightarrow{N \rightarrow \infty} \int_0^1 (x^2 + x(1 - 2x)) dx = \frac{1}{6}, \end{aligned}$$

which, jointly to (2.24), gives (2.23) for the irrational case.

Assume  $\alpha$  to be rational, say  $\alpha = \frac{p}{q}$  with  $p, q$  coprime numbers and  $q \geq 1$ . Let us define the function  $f : (x, y) \mapsto (x - \mathbb{1}_{x \geq 1-y})^2$  on  $[0, 1]^2$ , in order to write  $(\omega_j^{(N)})^2 = f(\{j\delta_N\}, \{j\alpha\})$  for all  $j$ . We slightly adapt the previous proof, starting again with the case  $\theta = 1$ .

We decompose into three parts

$$\begin{aligned} \sum_{j=1}^N \frac{1}{j} f(\{j\delta_N\}, \{j\alpha\}) &= \sum_{j=1}^{\lceil \frac{1}{\delta_N} \rceil - 1} \frac{1}{j} f(j\delta_N, \{j\alpha\}) + (A_N + B_N) \\ &\quad + \sum_{j=\lceil \frac{1}{\delta_N} \rceil}^{N-1} \frac{1}{j+1} (A_j + B_j) \end{aligned}$$

with for all  $j \geq 1$ ,

$$A_j := \frac{1}{j} \sum_{k=\lceil \frac{1}{\delta_N} \rceil}^{\lceil j\delta_N \rceil / \delta_N - 1} f(\{k\delta_N\}, \{k\alpha\})$$

and

$$B_j := \frac{1}{j} \sum_{k=\lceil j\delta_N \rceil / \delta_N}^j f(\{k\delta_N\}, \{k\alpha\}).$$

For the first part, noticing that for all  $x \in [0, \frac{1}{2}]$  and  $j \geq 1$  we have  $f(x, \{j\alpha\}) \leq f(x, \frac{q-1}{q})$ , then

$$\begin{aligned} 0 \leq \sum_{j=1}^{\lceil \frac{1}{\delta_N} \rceil - 1} \frac{1}{j} f(j\delta_N, \{j\alpha\}) &\leq \sum_{j=1}^{\lceil \frac{1}{2\delta_N} \rceil - 1} \frac{1}{j} f\left(j\delta_N, \frac{q-1}{q}\right) + \sum_{j=\lceil \frac{1}{2\delta_N} \rceil}^{\lceil \frac{1}{\delta_N} \rceil - 1} \frac{1}{j} \\ &\xrightarrow{N \rightarrow \infty} \int_0^{1/2} \frac{(x - \mathbb{1}_{x \geq \frac{1}{q}})^2}{x} dx + \int_{1/2}^1 \frac{1}{x} dx < +\infty. \end{aligned}$$

For the third part, since  $f$  is bounded by 1 on  $[0, 1]^2$ , it is easy to check that

$$\sum_{j=\lceil \frac{1}{\delta_N} \rceil}^{N-1} \frac{1}{j+1} B_j = \mathcal{O}(1). \text{ Moreover, } A_j \text{ for } j \geq 1 \text{ can be formulated as}$$

$$\begin{aligned} A_j &= \frac{1}{j} \sum_{l=1}^{\lfloor j\delta_N \rfloor - 1} \left( \sum_{k=\lceil \frac{l}{q\delta_N} \rceil}^{\lfloor \frac{1}{q} \lceil \frac{l+1}{\delta_N} \rceil \rfloor - 1} \sum_{m=0}^{q-1} f_m(\{(kq+m)\delta_N\}) \right) \\ &\quad + \sum_{r=\lceil \frac{l}{\delta_N} \rceil}^{q\lceil \frac{l}{q\delta_N} \rceil - 1} f_r(\{r\delta_N\}) + \sum_{r=q\lceil \frac{1}{q} \lceil \frac{l+1}{\delta_N} \rceil \rfloor}^{\lceil \frac{l+1}{\delta_N} \rceil - 1} f_r(\{r\delta_N\}) \end{aligned}$$

## 2. COUNTING FUNCTION OF EIGENVALUES AT MACROSCOPIC AND MESOSCOPIC SCALES

---

where  $f_n(x) := f(x, \{n\alpha\})$  for all  $n$ . This new expression of  $A_j$  is manageable. Indeed,

$$\begin{aligned} \frac{1}{j} \sum_{l=1}^{\lfloor j\delta_N \rfloor - 1} \left( \sum_{r=\lceil \frac{l}{\delta_N} \rceil}^{q\lceil \frac{l}{q\delta_N} \rceil - 1} f_r(\{r\delta_N\}) + \sum_{r=q\lceil \frac{1}{q} \lceil \frac{l+1}{\delta_N} \rceil \rceil}^{\lceil \frac{l+1}{\delta_N} \rceil - 1} f_r(\{r\delta_N\}) \right) &\leq \frac{1}{j} (\lfloor j\delta_N \rfloor - 1) (q + q) \\ &\leq 2q\delta_N, \end{aligned}$$

and for all  $l, m$ ,

$$\sum_{k=\lceil \frac{l}{q\delta_N} \rceil}^{\lfloor \frac{1}{q} \lceil \frac{l+1}{\delta_N} \rceil \rfloor - 1} f_m(\{(kq + m)\delta_N\}) = \sum_{k=0}^{\lfloor \frac{1}{q} \lceil \frac{l+1}{\delta_N} \rceil \rfloor - 1 - \lceil \frac{l}{q\delta_N} \rceil} f_m\left(\left(kq + m + q \left\{ \frac{-l}{q\delta_N} \right\}\right) \delta_N\right).$$

Let  $\tilde{A}_j := \frac{1}{j} (\lfloor j\delta_N \rfloor - 1) \sum_{m=0}^{q-1} \sum_{k=0}^{\lfloor \frac{1}{q\delta_N} \rfloor - 1} f_m((kq + m)\delta_N)$ . Then

$$\begin{aligned} |A_j - \tilde{A}_j| &\leq \frac{1}{j} \sum_{l=1}^{\lfloor j\delta_N \rfloor - 1} \sum_{m=0}^{q-1} \sum_{k=0}^{\lfloor \frac{1}{q\delta_N} \rfloor - 1} \left| f_m\left(\left(kq + m + q \left\{ \frac{-l}{q\delta_N} \right\}\right) \delta_N\right) - f_m((kq + m)\delta_N) \right| \\ &\quad + \mathcal{O}(\delta_N). \end{aligned}$$

For all  $0 \leq m \leq q-1$ , the function  $f_m$  is piecewise polynomial on  $[0, 1]$  (it has one finite discontinuity at point  $t = 1 - \{m\alpha\}$ ). Thus

$$\begin{aligned} &\left| f_m\left(\left(kq + m + q \left\{ \frac{-l}{q\delta_N} \right\}\right) \delta_N\right) - f_m((kq + m)\delta_N) \right| \\ &\leq q \left\{ \frac{-l}{q\delta_N} \right\} \delta_N \operatorname{ess\,sup}_{[0,1]} |f'_m| \\ &\quad + \mathbb{1}_{t \leq (kq + m + q \left\{ \frac{-l}{q\delta_N} \right\}) \delta_N < t + q \left\{ \frac{-l}{q\delta_N} \right\} \delta_N} \\ &\leq 2q\delta_N + \mathbb{1}_{t \leq (kq + m + q \left\{ \frac{-l}{q\delta_N} \right\}) \delta_N < t + q \left\{ \frac{-l}{q\delta_N} \right\} \delta_N}. \end{aligned}$$

At fixed  $l, m$ , it is easy to check that there is at most one  $k$  such that the indicator function equals 1. Hence

$$|A_j - \tilde{A}_j| \leq \frac{1}{j} \times j\delta_N \times q \times \left[ \frac{1}{q\delta_N} \times (2q\delta_N) + 1 \right] + \mathcal{O}(\delta_N) = \mathcal{O}(\delta_N)$$

and consequently

$$\left| \sum_{j=\lceil 1/\delta_N \rceil}^{N-1} \frac{A_j}{j+1} - \sum_{j=\lceil 1/\delta_N \rceil}^{N-1} \frac{\tilde{A}_j}{j+1} \right| = o(\log(N\delta_N)).$$

Moreover,

$$\begin{aligned} \sum_{j=\lceil 1/\delta_N \rceil}^{N-1} \frac{\tilde{A}_j}{j+1} &= \delta_N \sum_{m=0}^{q-1} \sum_{k=0}^{\lfloor \frac{1}{q\delta_N} \rfloor - 1} f_m((kq+m)\delta_N) \left( \sum_{j=\lceil 1/\delta_N \rceil}^{N-1} \frac{1}{j+1} + \mathcal{O}(1) \right) \\ &= \frac{1}{q} \sum_{m=0}^{q-1} \left( \int_0^1 f_m(x) dx + o(1) \right) (\log(N\delta_N) + \mathcal{O}(1)). \end{aligned}$$

Since  $p$  and  $q$  are coprime numbers, the numbers  $\{m\alpha\}$  cycle through some rearrangement of the numbers  $0, \frac{1}{q}, \dots, \frac{q-1}{q}$ , hence

$$\begin{aligned} \frac{1}{q} \sum_{m=0}^{q-1} \int_0^1 f_m(x) dx &= \frac{1}{q} \sum_{m=0}^{q-1} \int_0^1 \left( x - \mathbb{1}_{x \geq 1 - \frac{m}{q}} \right)^2 dx \\ &= \frac{1}{3} - \frac{1}{q} \sum_{m=1}^{q-1} \frac{m}{q} \left( 1 - \frac{m}{q} \right) \\ &= \frac{1}{6} + \frac{1}{6q^2} \end{aligned}$$

(indeed when  $q = 1$  the last equality is satisfied too).

Furthermore, note that  $|\tilde{A}_N| \leq \delta_N \sum_{m=0}^{q-1} \sum_{k=0}^{\lfloor \frac{1}{q\delta_N} \rfloor - 1} 1 \leq 1$  and then  $|A_N + B_N| = \mathcal{O}(1)$ .

Finally, putting it all together, we have

$$\sum_{j=1}^N \frac{(\omega_j^{(N)})^2}{j} = \left( \frac{1}{6} + \frac{1}{6q^2} \right) \log(N\delta_N) + o(\log(N\delta_N)) \quad (2.25)$$

and using the same argument as in the previous proof we extend (2.25) to all  $\theta > 0$ , which provides

$$\sum_{j=1}^N \frac{(\omega_j^{(N)})^2}{j} \Psi_N(j) = \left( \frac{1}{6} + \frac{1}{6q^2} \right) \log(N\delta_N) + o(\log(N\delta_N))$$

From (2.24) we deduce (2.23) for the rational case.  $\square$

### 2.4.2 Limiting normality for a single mesoscopic arc

**Theorem 2.4.4.** *Suppose that the sequence  $(\delta_N)$  satisfies* 
$$\begin{cases} \delta_N \xrightarrow{N \rightarrow \infty} 0 \\ N\delta_N \xrightarrow{N \rightarrow \infty} +\infty. \end{cases}$$

*Let  $\tilde{Y}_N^{I_N} := \frac{\tilde{X}_N^{I_N} - \mathbb{E}(\tilde{X}_N^{I_N})}{(\frac{\theta}{6} \log(N\delta_N))^{1/2}}$ . Then*

$$\tilde{Y}_N^{I_N} \xrightarrow{d} \mathcal{N}(0, 1).$$

*Proof.* This proof is very similar to the proof of Theorem 2.3.10, in the particular case

of one interval.

We introduce again the Poisson variables  $(W_j)$  from Feller Coupling. We denote for all  $j, p$ ,  $b_{j,p}^{(N)} := \mathbb{1}_{\phi_{j,p} > \{j\alpha_N\}} - \mathbb{1}_{\phi_{j,p} > \{j\beta_N\}} + \{j\alpha_N\} - \{j\beta_N\}$ , and  $T_N := \sum_{j=1}^N V_{N,j}$  with for all  $j$ ,

$$V_{N,j} := \frac{1}{\left(\frac{\theta}{6} \log(N\delta_N)\right)^{1/2}} \sum_{p=1}^{W_j} b_{j,p}^{(N)}.$$

Let  $\varepsilon > 0$ . Using Markov's inequality and Lemma 2.2.8,

$$\begin{aligned} \mathbb{P}\left(|\tilde{Y}_N^{I_N} - T_N| > \varepsilon\right) &\leq \frac{1}{\varepsilon \left(\frac{\theta}{6} \log(N\delta_N)\right)^{1/2}} \sum_{j=1}^N \mathbb{E} \sum_{p=1}^{|a_{N,j} - W_j|} |b_{j,p}^{(N)}| \\ &= \frac{1}{\varepsilon \left(\frac{\theta}{6} \log(N\delta_N)\right)^{1/2}} \times \mathcal{O}_\theta(1). \end{aligned}$$

Hence  $\tilde{Y}_N^{I_N} - T_N \xrightarrow{\mathbb{P}} 0$ . Moreover, for all  $j$ ,

$$\text{Var}(V_{N,j}) = \frac{1}{\frac{1}{6} \log(N\delta_N)} \frac{1}{j} \{j\delta_N\}(1 - \{j\delta_N\}).$$

In a similar way as in the proof of Theorem 2.3.10, we have

$$\sum_{j=1}^N \text{Var}(V_{N,j}) = \frac{6}{\log(N\delta_N)} \sum_{j=1}^N \frac{\{j\delta_N\}(1 - \{j\delta_N\})}{j} \xrightarrow{N \rightarrow \infty} 1$$

and

$$\sum_{j=1}^N \mathbb{E} \left( (V_{N,j})^2 \mathbb{1}_{|V_{N,j}| > \varepsilon} \right) \leq \frac{1}{\frac{\theta}{6} \log(N\delta_N)} \sum_{j=1}^N \mathbb{E} \left( W_j^2 \mathbb{1}_{W_j > \varepsilon \left(\frac{\theta}{6} \log(N\delta_N)\right)^{1/2}} \right) \xrightarrow{N \rightarrow \infty} 0.$$

Then  $(T_N)$  converges in distribution to  $\mathcal{N}(0, 1)$ . Slutsky's theorem ends the proof.  $\square$

**Theorem 2.4.5.** Suppose that  $I_N = \left(e^{2i\pi\alpha}, e^{2i\pi(\alpha+\delta_N)}\right]$  with  $\begin{cases} \delta_N \xrightarrow{N \rightarrow \infty} 0 \\ N\delta_N \xrightarrow{N \rightarrow \infty} +\infty. \end{cases}$

Let  $Y_N^{I_N} := \frac{X_N^{I_N} - \mathbb{E}(X_N^{I_N})}{(\theta c_2(\alpha) \log(N\delta_N))^{1/2}}$  where  $c_2(\alpha)$  is a constant defined by

$$c_2(\alpha) := \begin{cases} \frac{1}{6} & \text{if } \alpha \text{ is irrational} \\ \frac{1}{6} + \frac{1}{6q^2} & \text{if } \alpha = \frac{p}{q}, p, q \text{ coprime numbers, } q \geq 1. \end{cases}$$

Then

$$Y_N^{I_N} \xrightarrow{d} \mathcal{N}(0, 1).$$

*Proof.* This proof is very similar to the proof of Theorem 2.3.9, in the particular case of one interval. We introduce again the Poisson variables  $(W_j)$  from Feller Coupling. Let

$T_N := \sum_{j=1}^N V_{N,j}$  with for all  $j$ ,

$$V_{N,j} := \frac{\frac{\theta}{j} - W_j}{(\theta c_2(\alpha) \log(N\delta_N))^{1/2}} \omega_j^{(N)}.$$

Let  $\varepsilon > 0$ . Using Markov's inequality, Lemma 2.2.7 and Lemma 2.2.8,

$$\begin{aligned} & \mathbb{P}\left(|Y_N^{I_N} - T_N| > \varepsilon\right) \\ & \leq \frac{1}{\varepsilon (\theta c_2(\alpha) \log(N\delta_N))^{1/2}} \left[ \sum_{j=1}^N \frac{\theta}{j} |\Psi_N(j) - 1| + \sum_{j=1}^N \mathbb{E}|W_j - a_{N,j}| \right] \\ & = \frac{1}{\varepsilon (\theta c_2(\alpha) \log(N\delta_N))^{1/2}} \times \mathcal{O}_\theta(1), \end{aligned}$$

hence  $Y_N^{I_N} - T_N \xrightarrow{\mathbb{P}} 0$ . In a similar way as in the proof of Theorem 2.3.9, we have

$$\sum_{j=1}^N \text{Var}(V_{N,j}) = \frac{1}{c_2(\alpha) \log(N\delta_N)} \sum_{j=1}^N \frac{(\omega_j^{(N)})^2}{j} \xrightarrow{N \rightarrow \infty} 1$$

and

$$\begin{aligned} & \sum_{j=1}^N \mathbb{E}\left((V_{N,j})^2 \mathbb{1}_{|V_{N,j}| > \varepsilon}\right) \\ & \leq \frac{1}{\theta c_2(\alpha) \log(N\delta_N)} \sum_{j=1}^N \mathbb{E}\left(\left(W_j - \frac{\theta}{j}\right)^2 \mathbb{1}_{\left|W_j - \frac{\theta}{j}\right| > \varepsilon (\theta c_2(\alpha) \log(N\delta_N))^{1/2}}\right) \\ & \xrightarrow{N \rightarrow \infty} 0. \end{aligned}$$

Then  $(T_N)$  converges in distribution to  $\mathcal{N}(0, 1)$ . Slutsky's theorem ends the proof.  $\square$

*Remark 2.4.6.* For a finite number of mesoscopic arcs, say  $m$  arcs, with shrinking speeds  $\delta_N^{(k)}$ ,  $1 \leq k \leq m$ , it is reasonable to expect that some asymptotic results still occur. Indeed, the only point to overcome in the proof is the existence of non-diagonal terms in the covariance matrices  $D$  and  $\tilde{D}$ , whose good candidates would be for  $k \neq l$ :

$$D_{k,l} = \frac{1}{(c_2(\alpha_k) c_2(\alpha_l))^{1/2}} \lim_{N \rightarrow \infty} \frac{1}{\left(\log(N\delta_N^{(k)}) \log(N\delta_N^{(l)})\right)^{1/2}} \sum_{j=1}^N \frac{\omega_{j,k}^{(N)} \omega_{j,l}^{(N)}}{j} \quad (2.26)$$

and

$$\tilde{D}_{k,l} = 6 \lim_{N \rightarrow \infty} \frac{1}{\left(\log(N\delta_N^{(k)}) \log(N\delta_N^{(l)})\right)^{1/2}} \sum_{j=1}^N \frac{H_{j,k,l}^{(N)}}{j}, \quad (2.27)$$

where

$$\begin{cases} \omega_{j,k}^{(N)} := \{j(\alpha_k + \delta_N^{(k)})\} - \{j\alpha_k\} \\ H_{j,k,l} = \frac{1}{2}(h_j(\alpha_N^{(k)} - \beta_N^{(l)}) + h_j(\beta_N^{(k)} - \alpha_N^{(l)}) - h_j(\alpha_N^{(k)} - \alpha_N^{(l)}) - h_j(\beta_N^{(k)} - \beta_N^{(l)})), \end{cases}$$

$h_j(x) := \{jx\}(1 - \{jx\})$ . It is not clear these limits exist, since the formulas suggest a deep dependence on the way the  $m$  arcs overlap when  $N$  becomes large.

## 2.5 Spacing between eigenvalues

For all  $N \geq 1$ , denote by  $\mathcal{D}_N$  and  $d_N$  (resp.  $\tilde{\mathcal{D}}_N$  and  $\tilde{d}_N$ ) the largest and the smallest spacings between two consecutive distinct eigenangles of a random element from  $\mathfrak{S}_N$  (resp.  $S^1 \wr \mathfrak{S}_N$ ), where the permutations are picked under Ewens measure of parameter  $\theta > 0$ .

### 2.5.1 Largest spacing between two consecutive distinct eigenvalues

**Proposition 2.5.1.** *The sequences of random variables  $(n\mathcal{D}_n)_{n \geq 1}$ ,  $(n\tilde{\mathcal{D}}_n)_{n \geq 1}$ ,  $(\frac{1}{n\mathcal{D}_n})_{n \geq 1}$  and  $(\frac{1}{n\tilde{\mathcal{D}}_n})_{n \geq 1}$  are tight.*

*Remark 2.5.2.* Informally, this proposition involves that  $\mathcal{D}_n$  and  $\tilde{\mathcal{D}}_n$  have an order of magnitude of  $\frac{1}{n}$ .

*Proof.* Let  $n \geq 1$ . It is easy to check that

$$\frac{2\pi}{Z_n} \leq \mathcal{D}_n \leq \frac{2\pi}{L_{n,1}}$$

where  $L_{n,1}$  denotes the largest cycle length of the corresponding permutation and  $Z_n$  the number of distinct eigenvalues, using the pigeonhole principle for the first inequality. In particular, since  $Z_n \leq n$ ,

$$2\pi \leq n\mathcal{D}_n \leq \frac{2\pi}{\frac{1}{n}L_{n,1}}. \quad (2.28)$$

Obviously, the left-hand side of (2.28) provides the tightness of  $(\frac{1}{n\mathcal{D}_n})_{n \geq 1}$ . Moreover, it is well-known (see [4]) that  $(\frac{1}{n}L_{n,1})$  converges in distribution to the first coordinate of a Poisson-Dirichlet random vector of parameter  $\theta$ , which is almost surely finite and positive. Using the continuous mapping theorem, it follows that  $(\frac{2\pi}{\frac{1}{n}L_{n,1}})$  converges in distribution, and thus this sequence is tight. We deduce

$$\forall \varepsilon > 0, \quad \exists \eta_\varepsilon > 0, \quad \forall n \geq 1, \quad \mathbb{P}(n\mathcal{D}_n \leq \eta_\varepsilon) \geq \mathbb{P}\left(\frac{2\pi}{\frac{1}{n}L_{n,1}} \leq \eta_\varepsilon\right) \geq 1 - \varepsilon.$$

Now, for a random element of the wreath product  $S^1 \wr \mathfrak{S}_n$  related to the same permutation, the previous inequality holds, *i.e.*

$$\frac{2\pi}{n} \leq \tilde{\mathcal{D}}_n \leq \frac{2\pi}{L_{n,1}}. \quad (2.29)$$



Note that in this case the number of distinct eigenvalues is almost surely equal to  $n$ . The same reasoning as above applied to  $\tilde{\mathcal{D}}_n$  gives the claim.  $\square$

### 2.5.2 Smallest spacing between two consecutive distinct eigenvalues

**Proposition 2.5.3.** *The sequences of random variables  $(n^2 d_n)_{n \geq 1}$ ,  $(n^2 \tilde{d}_n)_{n \geq 1}$ ,  $(\frac{1}{n^2 d_n})_{n \geq 1}$  and  $(\frac{1}{n^2 \tilde{d}_n})_{n \geq 1}$  are tight.*

*Remark 2.5.4.* Informally, this proposition involves that  $d_n$  and  $\tilde{d}_n$  have an order of magnitude of  $\frac{1}{n^2}$ .

*Proof.* Let  $(a_{n,1}, \dots, a_{n,n})$  be the cycle structure and  $(A_{n,1}, \dots, A_{n,n})$  be the age-ordered list of cycle lengths (i.e. the vector of cycle lengths in order of appearance, that is to say the increasing order following the lowest element of each cycle) of the random  $n$ -permutation.

The smallest spacing  $d_n$  can be formulated as

$$d_n := \frac{2\pi}{\sup\{\text{lcm}(k, l) : 1 \leq k, l \leq n, a_{n,k}, a_{n,l} \geq 1\}} \quad (2.30)$$

On the one hand, trivially  $d_n \geq \frac{2\pi}{n^2}$  for all  $n$ , then  $(\frac{1}{n^2 d_n})_{n \geq 1}$  is tight.

On the other hand,

$$d_n \leq \frac{2\pi}{\text{lcm}(A_{n,1}, A_{n,2})} = \frac{2\pi \cdot \gcd(A_{n,1}, A_{n,2})}{A_{n,1} A_{n,2}}. \quad (2.31)$$

We are going to show that  $(\gcd(A_{n,1}, A_{n,2}))_{n \geq 1}$  is tight.

Let  $A$  be a positive integer.

$$\begin{aligned} \mathbb{P}(\gcd(A_{n,1}, A_{n,2}) \geq A) &= \sum_{j=A}^n \mathbb{P}(\gcd(A_{n,1}, A_{n,2}) = j) \\ &\leq \sum_{j=A}^n \mathbb{P}(j \text{ is a common divisor of } A_{n,1} \text{ and } A_{n,2}) \\ &= \sum_{j=A}^n \sum_{\substack{1 \leq k, l \\ k+l \leq \lfloor \frac{n}{j} \rfloor}} \mathbb{P}((A_{n,1}, A_{n,2}) = (jk, jl)). \end{aligned}$$

From a basic result in [4] we have, for all  $a_1, a_2 \geq 1$ ,

$$\mathbb{P}(A_{n,1} = a_1, A_{n,2} = a_2) = \frac{\theta^2 \Psi_n(a_1 + a_2)}{n(n - a_1)}.$$

Thus, for all  $j \geq 1$ ,

$$\sum_{\substack{1 \leq k, l \\ k+l \leq \lfloor \frac{n}{j} \rfloor}} \mathbb{P}((A_{n,1}, A_{n,2}) = (jk, jl)) = \frac{\theta^2}{n} \sum_{k=1}^{\lfloor \frac{n}{j} \rfloor - 1} \frac{1}{n - jk} \sum_{l=1}^{\lfloor \frac{n}{j} \rfloor - k} \Psi_n(j(k + l)).$$

If  $\theta \geq 1$ , then  $\Psi_n(m) \leq 1$  for all  $1 \leq m \leq n$ , and then

$$\frac{\theta^2}{n} \sum_{k=1}^{\lfloor \frac{n}{j} \rfloor - 1} \frac{1}{n - jk} \sum_{l=1}^{\lfloor \frac{n}{j} \rfloor - k} \Psi_n(j(k+l)) \leq \frac{\theta^2}{n} \sum_{k=1}^{\lfloor \frac{n}{j} \rfloor - 1} \frac{1}{n - jk} \left( \left\lfloor \frac{n}{j} \right\rfloor - k \right) \leq \frac{\theta^2}{j^2}.$$

If  $\theta < 1$ , then  $\Psi_n(m) \leq \Psi_n(m+1)$  for all  $1 \leq m \leq n-1$ , and then for all  $j \geq 1$  we have

$$\begin{aligned} \sum_{l=1}^{\lfloor \frac{n}{j} \rfloor - k - 1} \Psi_n(j(k+l)) &\leq \frac{1}{j} \sum_{l=1}^{\lfloor \frac{n}{j} \rfloor - k - 1} \sum_{p=j(l+1)-1}^{j(l+1)-1} \Psi_n(jk+p) \\ &\leq \Psi_n(jk) \frac{1}{j} \sum_{p=1}^{n-jk} \Psi_{n-jk}(p) \\ &= \Psi_n(jk) \frac{n-jk}{j\theta}, \end{aligned}$$

using Lemma 2.2.7 for the latest equality, and similarly

$$\sum_{k=1}^{\lfloor \frac{n}{j} \rfloor - 1} \Psi_n(jk) \leq \frac{1}{j} \sum_{k=1}^{\lfloor \frac{n}{j} \rfloor - 1} \sum_{p=jk}^{j(k+1)-1} \Psi_n(p) \leq \frac{n}{j\theta}.$$

Thus,

$$\begin{aligned} \frac{\theta^2}{n} \sum_{k=1}^{\lfloor \frac{n}{j} \rfloor - 1} \frac{1}{n - jk} \sum_{l=1}^{\lfloor \frac{n}{j} \rfloor - k} \Psi_n(j(k+l)) \\ \leq \frac{\theta^2}{n} \left[ \left( \frac{1}{j\theta} \times \frac{n}{j\theta} \right) + \sum_{k=1}^{\lfloor \frac{n}{j} \rfloor - 1} \frac{1}{n - jk} \Psi_n \left( j \left\lfloor \frac{n}{j} \right\rfloor \right) \right] \\ \leq \frac{1}{j^2} + \frac{\theta^2}{n} \Psi_n(n) \sum_{k=1}^{\lfloor \frac{n}{j} \rfloor - 1} \frac{1}{n - jk} \end{aligned}$$

with, using Holder inequality,

$$\begin{aligned} \sum_{k=1}^{\lfloor \frac{n}{j} \rfloor - 1} \frac{1}{n - jk} &\leq \frac{1}{j} \sum_{k=0}^{\lfloor \frac{n}{j} \rfloor - 1} \frac{1}{\left\lfloor \frac{n}{j} \right\rfloor - k} = \frac{1}{j} \sum_{k=1}^{\lfloor \frac{n}{j} \rfloor} \frac{1}{k} \leq \frac{1}{j} \left( \sum_{k=1}^{\lfloor \frac{n}{j} \rfloor} \frac{1}{k^{1/(1-\theta)}} \right)^{1-\theta} \left( \left\lfloor \frac{n}{j} \right\rfloor \right)^\theta \\ &= \mathcal{O}_\theta \left( \frac{n^\theta}{j^{1+\theta}} \right). \end{aligned}$$

Consequently, we deduce that for all  $\theta > 0$ ,

$$\begin{aligned} \mathbb{P}(\gcd(A_{n,1}, A_{n,2}) \geq A) &\leq (\theta^2 + 1) \sum_{j=A}^n \frac{1}{j^2} + \mathcal{O}_\theta \left( \sum_{j=A}^n \frac{\Psi_n(n)}{n^{1-\theta} j^{1+\theta}} \right) \\ &\leq (\theta^2 + 1) \sum_{j=A}^{+\infty} \frac{1}{j^2} + \mathcal{O}_\theta \left( \sum_{j=A}^{+\infty} \frac{1}{j^{1+\theta}} \right) \\ &\leq \varepsilon \end{aligned}$$

for all fixed real numbers  $\varepsilon > 0$  and  $A$  large enough depending on  $\theta$  and  $\varepsilon$ .

Finally, it is well-known (see again [4]) that as  $n \rightarrow \infty$

$$\frac{1}{n}(A_{n,1}, A_{n,2}, \dots, A_{n,n}) \xrightarrow{d} (G_1, G_2, \dots) \quad (2.32)$$

where  $(G_1, G_2, \dots)$  has GEM( $\theta$ ) distribution. Note that  $G_1$  and  $G_2$  are almost surely finite and positive. Since from (2.31)

$$2\pi \leq n^2 d_n \leq \frac{2\pi \gcd(A_{n,1}, A_{n,2})}{\frac{1}{n} A_{n,1} \cdot \frac{1}{n} A_{n,2}},$$

then the continuous mapping theorem gives the tightness of  $(n^2 d_n)_{n \geq 1}$ .

Now, for a random element of the wreath product  $S^1 \wr \mathfrak{S}_n$  related to the same permutation, the spacing is lower than the one for the permutation. Indeed, for all pair of cycle lengths  $(p, q)$ ,

- if  $p = q$ , then applying whatever rotations on the corresponding sets of eigenvalues (the  $p$ -th roots of unity) decreases the distance between two distinct eigenvalues.
- if  $p \neq q$ , we can suppose without loss of generality that  $p$  and  $q$  are coprime numbers (since the set of all eigenangles is periodic of period  $2\pi / \gcd(p, q)$ ). Let  $s \in (0, 1)$ . It is easy to check that the smallest spacing between two arguments of points which are respectively taken from  $\left\{ e^{\frac{2i\pi k}{p}}, 0 \leq k \leq p-1 \right\}$  and  $\left\{ e^{2i\pi(\frac{l}{q} + s)}, 0 \leq l \leq q-1 \right\}$  divided by  $2\pi$  is equal to

$$\min_{\substack{k \in \mathbb{Z} \\ l \in \mathbb{Z}}} \left| \frac{l}{q} + s - \frac{k}{p} \right| = \frac{1}{pq} \min_{\substack{k \in \mathbb{Z} \\ l \in \mathbb{Z}}} |lp - kq + spq|$$

and by Bézout we know that  $\{lp - kq : k, l \in \mathbb{Z}\} = \mathbb{Z}$ , thus

$$\min_{\substack{k \in \mathbb{Z} \\ l \in \mathbb{Z}}} |lp - kq + spq| = \min_{n \in \mathbb{Z}} |n + spq| = \min(\{spq\}, 1 - \{spq\}) \leq 1.$$

Consequently we have  $\tilde{d}_n \leq d_n$  and we deduce the tightness of  $(n^2 \tilde{d}_n)_{n \geq 1}$ .

It remains to show that  $(\frac{1}{n^2 \tilde{d}_n})_{n \geq 1}$  is tight. For this purpose, denote  $E_n$  the ensemble of couples of cycle lengths of the considered randomly chosen  $n$ -permutation. First

observe that for all uniform random variables  $U$  on  $[0, 1]$  and all non-zero integer  $n$ ,  $\{nU\}$  is uniform on  $[0, 1]$  and then  $\min(\{nU\}, 1 - \{nU\})$  is uniform on  $[0, 1/2]$ . Hence, using what we did above,

$$\tilde{d}_n \geq 2\pi \min_{(l_1, l_2) \in E_n} \frac{1}{l_1 l_2} V_{l_1, l_2} \quad (2.33)$$

where  $V_{j,k}$  are uniform random variables on  $[0, 1/2]$  (which are not independent when the indices overlap).

Now, conditionally to  $E_n = E$  where  $E$  is a possible ensemble of couples of lengths for the picked  $n$ -permutation, we have for all positive real numbers  $t < \frac{1}{2}$ ,

$$\begin{aligned} \mathbb{P} \left( \min_{(l_1, l_2) \in E_n} \frac{1}{l_1 l_2} V_{l_1, l_2} \leq \frac{t}{n^2} \mid E_n = E \right) &\leq \sum_{(l_1, l_2) \in E} \mathbb{P} \left( V_{l_1, l_2} \leq \frac{t l_1 l_2}{n^2} \right) \\ &\leq \sum_{(l_1, l_2) \in E} 2 \frac{t l_1 l_2}{n^2} = 2t \end{aligned}$$

using the union bound for the first inequality. Thus,

$$\mathbb{P} \left( \min_{(l_1, l_2) \in E_n} \frac{1}{l_1 l_2} V_{l_1, l_2} \leq \frac{t}{n^2} \right) \leq 2t$$

and, from (2.33),

$$\mathbb{P} \left( \frac{1}{n^2 \tilde{d}_n} > A \right) \leq \frac{1}{\pi A}$$

which gives the required tightness.  $\square$

## Appendix A

We give a proof of Lemma 2.2.8 using Cesàro means. Our proof provide a slightly better upper-bound than the one given in [4]. We begin with the following lemma:

**Lemma 2.5.5.** *For all  $1 \leq j \leq n$ ,*

$$\sum_{p=j}^{n-1} \frac{A^{\theta-1} p^{-j}}{p A_p^\theta} = \Psi_n(j) \left( \frac{1}{j} - \frac{1}{n} \right). \quad (2.34)$$

*Proof.* With the same notation as in the proof of Proposition 2.2.6, we observe that

$$\begin{aligned} J_n(Ms)_n &= \sum_{p=1}^{n-1} J_n M_{n,p} s_p + s_n = \sum_{p=1}^{n-1} \frac{\theta}{\theta + p} \frac{1}{p} \sum_{j=1}^p \Psi_p(j) w_j + \frac{1}{n} \sum_{j=1}^n \Psi_n(j) w_j \\ &= \sum_{j=1}^{n-1} w_j \left[ \sum_{p=j}^{n-1} \frac{\theta}{\theta + p} \frac{\Psi_p(j)}{p} + \frac{\Psi_n(j)}{n} \right] + \frac{\Psi_n(n)}{n} w_n \end{aligned}$$

and

$$J_n t_n = \sum_{j=1}^{n-1} w_j \frac{\Psi_n(j)}{j} + \frac{\Psi_n(n)}{n} w_n.$$

Therefore, by identification, for all  $j \in \llbracket 1, n-1 \rrbracket$ ,

$$\frac{\Psi_n(j)}{j} = \sum_{p=j}^{n-1} \frac{\theta}{\theta+p} \frac{\Psi_p(j)}{p} + \frac{\Psi_n(j)}{n}.$$

Finally, using the definition of Cesàro numbers it is clear that  $\frac{\theta}{\theta+p} \frac{\Psi_p(j)}{p} = \frac{A_{p-j}^{\theta-1}}{pA_p^\theta}$  for all  $1 \leq j \leq p$ .  $\square$

We are ready to prove Lemma 2.2.8.

We introduce independent Bernoulli variables  $\xi_r$ ,  $r \geq 1$ , defined as

$$\mathbb{P}(\xi_r = 1) = \frac{\theta}{\theta + r - 1} \quad , \quad \mathbb{P}(\xi_r = 0) = \frac{r-1}{\theta + r - 1}.$$

The Feller Coupling characterizes the variables  $a_{n,j}$  and  $W_j$  on the same probability space in function of the  $\xi_r$ , by the following equalities: For all  $1 \leq j \leq n$ ,

$$\begin{aligned} a_{n,j} &= \#\{j - \text{spacings between two consecutive 1 in the word } (1 \ \xi_2 \ \cdots \ \xi_n \ 1)\} \\ &= \sum_{k=1}^{n-j} \xi_k (1 - \xi_{k+1}) \cdots (1 - \xi_{k+j-1}) \xi_{k+j} + \xi_{n-j+1} (1 - \xi_{n-j+1}) \cdots (1 - \xi_n), \end{aligned}$$

and for all  $j \in \mathbb{N}^*$ ,

$$W_j = \sum_{k=1}^{+\infty} \xi_k (1 - \xi_{k+1}) \cdots (1 - \xi_{k+j-1}) \xi_{k+j}.$$

To begin with, it is easy to notice that

$$|a_{n,j} - W_j| \leq \mathbb{1}_{J_n=j} + W_{j,n} + \mathbb{1}_{J_n+K_n=j+1}$$

where

$$\begin{cases} J_n := \min\{k \geq 1 : \xi_{n-k+1} = 1\} \\ K_n := \min\{k \geq 1 : \xi_{n+k} = 1\} \\ W_{j,n} := \sum_{k=n+1}^{+\infty} \xi_k (1 - \xi_{k+1}) \cdots (1 - \xi_{k+j-1}) \xi_{k+j} \end{cases}$$

therefore

$$\mathbb{E}|a_{n,j} - W_j| \leq \mathbb{P}(J_n = j) + \mathbb{E}(W_{j,n}) + \mathbb{P}(J_n + K_n = j+1).$$

We look separately at the three right-hand side terms in this inequality.

$$\mathbb{P}(J_n = j) = \frac{\theta}{\theta + n - j} \frac{n - j + 1}{\theta + n - j + 1} \cdots \frac{n - 1}{\theta + n - 1} = \frac{\theta}{n} \Psi_n(j).$$

Hence,  $\sum_{j=1}^n \mathbb{P}(J_n = j) = \frac{\theta}{n} \sum_{j=1}^n \Psi_n(j) = 1$ .

$$\begin{aligned} \mathbb{E}(W_{j,n}) &= \sum_{k=n+1}^{+\infty} \frac{\theta}{\theta + k - 1} \frac{k}{\theta + k} \cdots \frac{k + j - 2}{\theta + k + j - 2} \frac{\theta}{\theta + k + j - 1} \\ &= \sum_{p=n+j+1}^{+\infty} \frac{\theta}{\theta + p - j - 1} \frac{p - j}{\theta + p - j} \cdots \frac{p - 2}{\theta + p - 2} \frac{\theta}{\theta + p - 1}. \end{aligned}$$

$$\begin{aligned} \mathbb{P}(J_n + K_n = j + 1) &= \sum_{k=1}^j \frac{\theta}{\theta + n + k - j - 1} \frac{n + k - j}{\theta + n + k - j} \cdots \frac{n + k - 2}{\theta + n + k - 2} \frac{\theta}{\theta + n + k - 1} \\ &= \sum_{p=n+1}^{n+j} \frac{\theta}{\theta + p - j - 1} \frac{p - j}{\theta + p - j} \cdots \frac{p - 2}{\theta + p - 2} \frac{\theta}{\theta + p - 1}. \end{aligned}$$

In particular,

$$\begin{aligned} \mathbb{E}(W_{j,n}) + \mathbb{P}(J_n + K_n = j + 1) &= \sum_{p=n+1}^{+\infty} \frac{\theta}{\theta + p - j - 1} \frac{p - j}{\theta + p - j} \cdots \frac{p - 2}{\theta + p - 2} \frac{\theta}{\theta + p - 1} \\ &= \theta \sum_{p=n}^{+\infty} \frac{1}{p} \frac{A_{p-j}^{\theta-1}}{A_p^\theta} = \theta \left[ \sum_{p=j}^{+\infty} \frac{1}{p} \frac{A_{p-j}^{\theta-1}}{A_p^\theta} - \sum_{p=j}^{n-1} \frac{1}{p} \frac{A_{p-j}^{\theta-1}}{A_p^\theta} \right]. \end{aligned}$$

Since, by Lemma 2.5.5,

$$\sum_{p=j}^{n-1} \frac{1}{p} \frac{A_{p-j}^{\theta-1}}{A_p^\theta} = \Psi_n(j) \left( \frac{1}{j} - \frac{1}{n} \right) \xrightarrow{n \rightarrow \infty} \frac{1}{j}$$

it follows

$$\mathbb{E}(W_{j,n}) + \mathbb{P}(J_n + K_n = j + 1) = \theta \left[ \frac{1}{j} - \Psi_n(j) \left( \frac{1}{j} - \frac{1}{n} \right) \right].$$

Then,

$$\begin{aligned} \sum_{j=1}^n \mathbb{E}(W_{j,n}) + \mathbb{P}(J_n + K_n = j + 1) &= \theta \sum_{j=1}^n \frac{1}{j} - \theta \sum_{j=1}^n \frac{\Psi_n(j)}{j} + \frac{\theta}{n} \sum_{j=1}^n \Psi_n(j) \\ &= \theta \sum_{j=1}^n \frac{1}{j} - \theta \sum_{j=1}^n \frac{1}{\theta + j - 1} + 1 \\ &= \theta \sum_{j=1}^n \frac{\theta - 1}{j(\theta + j - 1)} + 1 \end{aligned}$$

using Lemma 2.2.7 for the second equality. We deduce

$$\begin{aligned} \mathbb{E} \left( \sum_{j=1}^n |a_{n,j} - W_j| \right) &\leq 2 + \theta(\theta - 1) \sum_{j=1}^n \frac{1}{j(\theta + j - 1)} \\ &\leq 1 + \theta \mathbf{1}_{0 < \theta < 1} + (1 + \theta(\gamma + \psi(\theta))) \mathbf{1}_{\theta \geq 1} \end{aligned} \quad (2.35)$$

where  $\gamma$  is the Euler-Mascheroni constant, and  $\psi$  is the digamma function. As a particular consequence,

$$\limsup_{\theta \rightarrow 0^+} \sup_{n \geq 1} \mathbb{E} \left( \sum_{j=1}^n |a_{n,j} - W_j| \right) \leq 1. \quad (2.36)$$

## Appendix B

The following proposition gives the whole possible values for the constant  $c_2$  appearing in Proposition 2.3.2, in function of  $\alpha$  and  $\beta$ .

**Proposition 2.5.6.** *Let  $p, r$  be integers and  $q, s$  positive integers with  $p$  and  $q$  relatively prime,  $r$  and  $s$  relatively prime,  $r \neq 0$ . Then*

$$c_2 = \begin{cases} \frac{1}{6} & \text{if } \alpha \text{ and } \beta \text{ are irrational and} \\ & \text{linearly independent over } \mathbb{Q} \\ \frac{1}{6} + \frac{1}{6q^2} & \text{if } \alpha = \frac{p}{q} \text{ and } \beta \text{ is irrational} \\ \frac{1}{6} + \frac{1}{6s^2} & \text{if } \alpha \text{ is irrational and } \beta = \frac{r}{s} \\ \frac{(2q-1)(q-1)}{6q^2} + \frac{(2s-1)(s-1)}{6s^2} - \frac{2}{qs} \sum_{j=1}^{qs} \left\{ \frac{jp}{q} \right\} \left\{ \frac{jr}{s} \right\} & \text{if } \alpha = \frac{p}{q} \text{ and } \beta = \frac{r}{s} \\ \frac{1}{6} - \frac{\gcd(s,q)^2}{6srq^2} & \text{if } \alpha \text{ is irrational and} \\ & \beta = \frac{p}{q} + \frac{r}{s}\alpha. \end{cases}$$

The four first cases have been shown by Wieand, as well as the last case for  $s = 1$  by the same author in [87]. We complete her work in this appendix, treating the case  $s$  arbitrary.

Therefore, suppose  $\alpha$  and  $\beta$  irrational numbers which are linearly dependent over  $\mathbb{Q}$ , say  $\beta = \frac{p}{q} + \frac{r}{s}\alpha$ , with  $s \geq 2$ .

Let us recall that  $c_2 := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n (\{j\beta\} - \{j\alpha\})^2$ . Since  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \{jx\}^2$  is easy to compute for all real numbers  $x$ , it remains to study what was denoted by  $s_3$  in [87], defined as

$$s_3(\alpha, \beta) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \{j\alpha\} \{j\beta\}. \quad (2.37)$$

We first assume that we have  $r > 0$ . We are going to proceed in two steps.

- Computation of  $s_3(\alpha, \frac{r}{s}\alpha)$  :

Before starting the calculation, it is good to notice that if  $\psi := \frac{\alpha}{s}$  then  $\psi$  is still irrational and consequently the sequence  $(n\psi)_{n \in \mathbb{N}^*}$  is still equidistributed.

Thus  $s_3(s\psi, r\psi) = s_3(s\alpha, r\alpha)$ , i.e.  $s_3(\alpha, \frac{r}{s}\alpha) = s_3(s\alpha, r\alpha)$ . Indeed, it is a direct consequence of Theorem 9 given in [87] that we recall here:

**Theorem 2.5.7.** *Let  $f$  be a Riemann integrable function on  $[0, 1]$ , and let  $t \in \mathbb{R} \setminus \mathbb{Q}$ . Then for all real numbers  $b$ ,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n f(\{jt + b\}) = \int_0^1 f(x) dx.$$

(We take  $b = 0$  and  $f(x) = \{sx\}\{rx\}$ .)

The expression of  $s_3(s\alpha, r\alpha)$  is more convenient to handle in practice. We decompose for  $j \geq 1$ ,

$$\begin{aligned} \{js\alpha\}\{jr\alpha\} &= (s\{j\alpha\} - \lfloor s\{j\alpha\} \rfloor)(r\{j\alpha\} - \lfloor r\{j\alpha\} \rfloor) \\ &= rs\{j\alpha\}^2 - r \sum_{l=1}^{s-1} \{j\alpha\} \mathbb{1}_{\{j\alpha\} \geq \frac{l}{s}} - s \sum_{k=1}^{r-1} \{j\alpha\} \mathbb{1}_{\{j\alpha\} \geq \frac{k}{r}} \\ &\quad + \sum_{\substack{1 \leq k \leq r-1 \\ 1 \leq l \leq s-1}} \mathbb{1}_{\{j\alpha\} \geq \max(\frac{l}{s}, \frac{k}{r})} \end{aligned}$$

The three first terms will not pose any difficulty since the limits of their respective means have already been evaluated in [87]. The novelty holds in the fourth term. For  $k \in \llbracket 1, r-1 \rrbracket$  and  $l \in \llbracket 1, s-1 \rrbracket$ , let  $f_{k,l}(x) := \mathbb{1}_{x \geq \max(\frac{l}{s}, \frac{k}{r})}$ . These functions are clearly Riemann-integrable on  $[0, 1]$ , thus

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n \sum_{\substack{1 \leq k \leq r-1 \\ 1 \leq l \leq s-1}} \mathbb{1}_{\{j\alpha\} \geq \max(\frac{l}{s}, \frac{k}{r})} &\xrightarrow{n \rightarrow \infty} \sum_{\substack{1 \leq k \leq r-1 \\ 1 \leq l \leq s-1}} \int_0^1 f_{k,l}(x) dx \\ &= (r-1)(s-1) - \sum_{\substack{1 \leq k \leq r-1 \\ 1 \leq l \leq s-1}} \max\left(\frac{l}{s}, \frac{k}{r}\right) \end{aligned}$$

with

$$\begin{aligned} &\sum_{\substack{1 \leq k \leq r-1 \\ 1 \leq l \leq s-1}} \max\left(\frac{l}{s}, \frac{k}{r}\right) \\ &= \sum_{k=1}^{r-1} \left( \left( \sum_{l=1}^{\lfloor \frac{sk}{r} \rfloor} \frac{k}{r} \right) + \left( \sum_{l=\lfloor \frac{sk}{r} \rfloor + 1}^{s-1} \frac{l}{s} \right) \right) \\ &= \sum_{k=1}^{r-1} \left( \left( \frac{k}{r} \left\lfloor \frac{sk}{r} \right\rfloor \right) + \left( \frac{s-1}{2} - \frac{\lfloor \frac{sk}{r} \rfloor (\lfloor \frac{sk}{r} \rfloor + 1)}{2s} \right) \right) \\ &= \frac{(r-1)(s-1)}{2} + \frac{1}{2s} \sum_{k=1}^{r-1} \left( \frac{sk}{r} \right)^2 - \frac{sk}{r} - \left\{ \frac{sk}{r} \right\}^2 + \left\{ \frac{sk}{r} \right\} \\ &= \frac{(r-1)(s-1)}{2} + \frac{1}{2s} \left( \frac{s^2-1}{r^2} \frac{r(r-1)(2r-1)}{6} - \frac{s-1}{r} \frac{r(r-1)}{2} \right). \end{aligned}$$



Once expanded, we get

$$(r-1)(s-1) - \sum_{\substack{1 \leq k \leq r-1 \\ 1 \leq l \leq s-1}} \max\left(\frac{l}{s}, \frac{k}{r}\right) = \frac{1}{4} + \frac{rs}{3} - \frac{r}{4} - \frac{s}{4} - \frac{s}{12r} - \frac{r}{12s} + \frac{1}{12sr}.$$

We deduce

$$\begin{aligned} s_3(s\alpha, r\alpha) &= \frac{rs}{3} - r\left(\frac{s}{3} - \frac{1}{4} - \frac{1}{12s}\right) - s\left(\frac{r}{3} - \frac{1}{4} - \frac{1}{12r}\right) \\ &\quad + \left(\frac{1}{4} + \frac{rs}{3} - \frac{r}{4} - \frac{s}{4} - \frac{s}{12r} - \frac{r}{12s} + \frac{1}{12sr}\right) \\ &= \frac{1}{4} + \frac{1}{12sr} \end{aligned}$$

• Computation of  $s_3(\alpha, \frac{p}{q} + \frac{r}{s}\alpha)$  :

First of all, the particular case  $q = 1$  is trivial since  $s_3(\alpha, p + \frac{r}{s}\alpha) = s_3(\alpha, \frac{r}{s}\alpha)$  and then

$c_2 = \frac{2}{3} - 2\left(\frac{1}{4} + \frac{1}{12sr}\right) = \frac{1}{6} - \frac{\gcd(s,1)^2}{6sr \times 1^2}$ . Thus we assume in all the following that  $q \geq 2$ .

In the same manner as at the beginning of the first step, we notice that  $s_3(\alpha, \frac{p}{q} + \frac{r}{s}\alpha) = s_3(s\alpha, \frac{p}{q} + r\alpha)$  (using Theorem 2.5.7 with  $b = \frac{p}{q}$  and for  $f$  a complicated expression that we do not precise here, which is bounded and piecewise continuous so Riemann-integrable).

For all  $j \geq 1$ ,

$$\{js\alpha\}\{j\frac{p}{q} + jr\alpha\} = \{js\alpha\}\{j\frac{p}{q}\} + \{js\alpha\}\{jr\alpha\} - \{js\alpha\}\mathbb{1}_{\{j\frac{p}{q}\} + \{jr\alpha\} \geq 1}.$$

The mean of the first term tends to  $s_3(s\alpha, \frac{p}{q}) = \frac{1}{4} - \frac{1}{4q}$ . The mean of the second term tends to  $s_3(s\alpha, r\alpha) = \frac{1}{4} + \frac{1}{12sr}$  by the first step. It remains to study the third term. In a similar way to what is done in [87], it is easy to check that

$$\frac{1}{q \left\lfloor \frac{n}{q} \right\rfloor} \sum_{j=1}^{\left\lfloor \frac{n}{q} \right\rfloor} \{js\alpha\} \mathbb{1}_{\{j\frac{p}{q}\} + \{jr\alpha\} \geq 1} \underset{n \rightarrow \infty}{=} \frac{1}{n} \sum_{j=1}^n \{js\alpha\} \mathbb{1}_{\{j\frac{p}{q}\} + \{jr\alpha\} \geq 1} + o(1).$$

Let  $n \in \mathbb{N}^*$ . We use the periodicity of the sequence  $(\{\frac{kp}{q}\})$  to write that the left-hand side expression is equal to

$$\begin{aligned} &\frac{1}{q \left\lfloor \frac{n}{q} \right\rfloor} \sum_{j=0}^{\left\lfloor \frac{n}{q} \right\rfloor - 1} \sum_{k=1}^q \{(k+jq)s\alpha\} \mathbb{1}_{\{(k+jq)r\alpha\} \geq 1 - \{\frac{kp}{q}\}} \\ &= s \cdot \frac{1}{q} \sum_{k=1}^q \frac{1}{\left\lfloor \frac{n}{q} \right\rfloor} \sum_{j=0}^{\left\lfloor \frac{n}{q} \right\rfloor - 1} \{(k+jq)\alpha\} \mathbb{1}_{\{(k+jq)r\alpha\} \geq 1 - \{\frac{kp}{q}\}} \\ &\quad - \frac{1}{q} \sum_{k=1}^q \frac{1}{\left\lfloor \frac{n}{q} \right\rfloor} \sum_{j=0}^{\left\lfloor \frac{n}{q} \right\rfloor - 1} \sum_{m=1}^{s-1} \mathbb{1}_{\{(k+jq)\alpha\} \geq \frac{m}{s}} \cap \left( \{(k+jq)r\alpha\} \geq 1 - \{\frac{kp}{q}\} \right) \end{aligned}$$

## 2. COUNTING FUNCTION OF EIGENVALUES AT MACROSCOPIC AND MESOSCOPIC SCALES

---

with, (see [87]),

$$\frac{1}{q} \sum_{k=1}^q \frac{1}{\left\lfloor \frac{n}{q} \right\rfloor} \sum_{j=0}^{\left\lfloor \frac{n}{q} \right\rfloor - 1} \{(k+jq)\alpha\} \mathbb{1}_{\{(k+jq)r\alpha\} \geq 1 - \left\{ \frac{kp}{q} \right\}} \xrightarrow{n \rightarrow \infty} \frac{1}{4} + \frac{1}{12r} - \frac{1}{4q} - \frac{1}{12rq^2}.$$

For  $k \in \llbracket 1, q \rrbracket$  and  $m \in \llbracket 1, s-1 \rrbracket$ , let

$$f_{k,m}(x) = \mathbb{1}_{\left(x \geq \frac{m}{s}\right) \cap \left(rx - \sum_{l=1}^{r-1} \mathbb{1}_{x \geq \frac{l}{r}} \geq 1 - \left\{ \frac{kp}{q} \right\}\right)}.$$

We compute its integral between 0 and 1:

$$\begin{aligned} \int_0^1 f_{k,m}(x) dx &= \sum_{l=0}^{r-1} \int_{\frac{1 - \left\{ \frac{kp}{q} \right\} + l}{r}}^{\frac{l+1}{r}} \mathbb{1}_{\left(x \geq \frac{m}{s}\right)} dx \\ &= 0 + \left( \frac{\left\lfloor \frac{rm}{s} \right\rfloor + 1}{r} - \max \left( \frac{m}{s}, \frac{1 - \left\{ \frac{kp}{q} \right\} + \left\lfloor \frac{rm}{s} \right\rfloor}{r} \right) \right) \\ &\quad + \sum_{l=\left\lfloor \frac{rm}{s} \right\rfloor + 1}^{r-1} \frac{l+1}{r} - \frac{1 - \left\{ \frac{kp}{q} \right\} + l}{r} \\ &= \frac{1 - \left\{ \frac{rm}{s} \right\}}{r} - \frac{1 - \left\{ \frac{kp}{q} \right\} - \left\{ \frac{rm}{s} \right\}}{r} \mathbb{1}_{1 - \left\{ \frac{kp}{q} \right\} - \left\{ \frac{rm}{s} \right\} \geq 0} \\ &\quad + \left\{ \frac{kp}{q} \right\} \left( 1 - \frac{1}{r} - \frac{m}{s} + \frac{1}{r} \left\{ \frac{rm}{s} \right\} \right) \\ &= \frac{1 - \left\{ \frac{kp}{q} \right\} - \left\{ \frac{rm}{s} \right\}}{r} \mathbb{1}_{1 - \left\{ \frac{kp}{q} \right\} - \left\{ \frac{rm}{s} \right\} < 0} \\ &\quad + \left\{ \frac{kp}{q} \right\} \left( 1 - \frac{m}{s} + \frac{1}{r} \left\{ \frac{rm}{s} \right\} \right). \end{aligned}$$

Thus,

$$\begin{aligned} \sum_{k=1}^q \sum_{m=1}^{s-1} \int_0^1 f_{k,m}(x) dx &= - \sum_{k=1}^q \left\{ \frac{kp}{q} \right\} + \sum_{k=1}^q \sum_{m=0}^{s-1} \int_0^1 f_{k,m}(x) dx \\ &= - \sum_{k=0}^{q-1} \frac{k}{q} + \sum_{k=0}^{q-1} \sum_{m=0}^{s-1} \frac{1 - \frac{k}{q} - \frac{m}{s}}{r} \mathbb{1}_{\frac{m}{s} > 1 - \frac{k}{q}} \\ &\quad + \frac{k}{q} \left( 1 - \frac{m}{s} + \frac{m}{sr} \right) \\ &= \frac{q-1}{2} \left( s-1 - \frac{s-1}{2} + \frac{s-1}{2r} \right) \\ &\quad + \sum_{k=0}^{q-1} \sum_{m=0}^{s-1} \frac{1 - \frac{k}{q} - \frac{m}{s}}{r} \mathbb{1}_{\frac{m}{s} > 1 - \frac{k}{q}} \end{aligned}$$

where we get the second equality noting that the numbers  $\left\{\frac{kp}{q}\right\}$  (resp.  $\left\{\frac{mr}{s}\right\}$ ) cycle through some rearrangement of the numbers  $0, \frac{1}{q}, \dots, \frac{q-1}{q}$  (resp.  $0, \frac{1}{s}, \dots, \frac{s-1}{s}$ ).

Now, it just remains to compute the second term in the last expression, which is equal to

$$\frac{1}{r} \sum_{k=\lfloor \frac{q}{s} \rfloor + 1}^{q-1} \sum_{m=\lfloor s - \frac{ks}{q} \rfloor + 1}^{s-1} \left(1 - \frac{k}{q} - \frac{m}{s}\right).$$

Note that the condition  $\lfloor \frac{q}{s} \rfloor + 1 \leq q-1$  is always satisfied when  $q \geq 3$  since  $s \geq 2 > \frac{3}{2} \geq \frac{q}{q-1}$ . When  $q = 2$ , it only fails for  $s = 2$  (in this case we have an empty sum). For now assume we are not in the case  $q = s = 2$ . We will easily treat this remaining case at the end. We have:

$$\begin{aligned} \sum_{k=\lfloor \frac{q}{s} \rfloor + 1}^{q-1} \sum_{m=\lfloor s - \frac{ks}{q} \rfloor + 1}^{s-1} \left(1 - \frac{k}{q} - \frac{m}{s}\right) &= \sum_{k=\lfloor \frac{q}{s} \rfloor + 1}^{q-1} \left( (s-1) - \left\lfloor s - \frac{ks}{q} \right\rfloor \right) \left(1 - \frac{k}{q}\right) \\ &\quad - \frac{1}{s} \sum_{m=\lfloor s - \frac{ks}{q} \rfloor + 1}^{s-1} m \\ &= \sum_{k=\lfloor \frac{q}{s} \rfloor + 1}^{q-1} \frac{1}{2} \left( \frac{k}{q} - s \left( \frac{k}{q} \right)^2 \right) \\ &\quad + \frac{1}{2s} \left( \left\{ \frac{-ks}{q} \right\}^2 - \left\{ \frac{-ks}{q} \right\} \right). \end{aligned}$$

In addition, using the facts that for all  $x \in \mathbb{R}$ ,  $\{-x\}(1 - \{-x\}) = \{x\}(1 - \{x\})$ , and for all  $k \in \llbracket 0, \lfloor \frac{q}{s} \rfloor \rrbracket$ ,  $0 \leq \frac{ks}{q} < 1$ , it comes

$$\begin{aligned} \sum_{k=0}^{\lfloor \frac{q}{s} \rfloor} \frac{1}{2} \left( \frac{k}{q} - s \left( \frac{k}{q} \right)^2 \right) + \frac{1}{2s} \left( \left\{ \frac{-ks}{q} \right\}^2 - \left\{ \frac{-ks}{q} \right\} \right) \\ = \sum_{k=0}^{\lfloor \frac{q}{s} \rfloor} \frac{1}{2} \left( \frac{k}{q} - s \left( \frac{k}{q} \right)^2 \right) + \frac{1}{2s} \left( \left( \frac{ks}{q} \right)^2 - \frac{ks}{q} \right) = 0. \end{aligned}$$

Furthermore, denoting  $d := \gcd(s, q)$  and  $s = d\tilde{s}$ ,  $q = d\tilde{q}$ ,

$$\begin{aligned} \sum_{k=0}^{q-1} \left\{ \frac{ks}{q} \right\} \left(1 - \left\{ \frac{ks}{q} \right\}\right) &= d \sum_{k=0}^{\tilde{q}-1} \left\{ \frac{k\tilde{s}}{\tilde{q}} \right\} \left(1 - \left\{ \frac{k\tilde{s}}{\tilde{q}} \right\}\right) \\ &= d \sum_{k=0}^{\tilde{q}-1} \frac{k}{\tilde{q}} \left(1 - \frac{k}{\tilde{q}}\right) \\ &= \frac{q}{6} - \frac{d^2}{6q}. \end{aligned}$$

## 2. COUNTING FUNCTION OF EIGENVALUES AT MACROSCOPIC AND MESOSCOPIC SCALES

---

Thus,

$$\sum_{k=\lfloor \frac{q}{s} \rfloor + 1}^{q-1} \sum_{m=\lfloor s - \frac{ks}{q} \rfloor + 1}^{s-1} \left( 1 - \frac{k}{q} - \frac{m}{s} \right) = -\frac{q}{12s} + \frac{d^2}{12qs} + \frac{1}{2} \sum_{k=0}^{q-1} \left( \frac{k}{q} - s \left( \frac{k}{q} \right)^2 \right).$$

Expanding the expressions it follows

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n \{js\alpha\} \mathbb{1}_{\{j\frac{p}{q}\} + \{jr\alpha\} \geq 1} &\xrightarrow{n \rightarrow \infty} s \left( \frac{1}{4} + \frac{1}{12r} - \frac{1}{4q} - \frac{1}{12rq^2} \right) \\ &\quad - \frac{1}{q} \left( \frac{1}{4} + \frac{qs}{4} - \frac{q}{4} + \frac{qs}{4r} - \frac{q}{4r} - \frac{s}{4} - \frac{s}{4r} + \frac{1}{4r} \right) \\ &\quad - \frac{1}{qr} \left( -\frac{q}{12s} + \frac{d^2}{12sq} - \frac{qs}{6} + \frac{s}{4} - \frac{s}{12q} + \frac{q}{4} - \frac{1}{4} \right) \\ &= \frac{1}{4} - \frac{1}{4q} + \frac{1}{12sr} - \frac{d^2}{12srq^2} \end{aligned}$$

Consequently,

$$\begin{aligned} s_3 \left( \alpha, \frac{p}{q} + \frac{r}{s} \alpha \right) &= \left( \frac{1}{4} - \frac{1}{4q} \right) + \left( \frac{1}{4} + \frac{1}{12sr} \right) - \left( \frac{1}{4} - \frac{1}{4q} + \frac{1}{12sr} - \frac{d^2}{12srq^2} \right) \\ &= \frac{1}{4} + \frac{d^2}{12srq^2}. \end{aligned}$$

It is easy to check that this formula remains true for  $q = s = 2$ , since in this case we have

$$\sum_{k=1}^q \sum_{m=1}^{s-1} \int_0^1 f_{k,m}(x) dx = \frac{1}{4} + \frac{1}{4r}$$

and thus

$$\begin{aligned} s_3 \left( \alpha, \frac{p}{q} + \frac{r}{s} \alpha \right) &= \left( \frac{1}{4} - \frac{1}{4q} \right) + \left( \frac{1}{4} + \frac{1}{12sr} \right) \\ &\quad - s \left( \frac{1}{4} + \frac{1}{12sr} - \frac{1}{4q} - \frac{1}{12rq^2} \right) + \frac{1}{q} \left( \frac{1}{4} + \frac{1}{4r} \right) \\ &= \frac{1}{4} + \frac{1}{24r} = \frac{1}{4} + \frac{d^2}{12srq^2}. \end{aligned}$$

Finally, we assume  $r < 0$ .

As we have seen above, the result of  $s_3 \left( \alpha, \frac{p}{q} + \frac{r}{s} \alpha \right)$  is not depending on the choice of

the irrational  $\alpha$ ; thus, one can replace it by  $(-s)\alpha$ , and then

$$\begin{aligned}
s_3\left(\alpha, \frac{p}{q} + \frac{r}{s}\alpha\right) &= s_3\left(s(-\alpha), \frac{p}{q} + |r|\alpha\right) \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n (1 - \{js\alpha\}) \left\{j \left(\frac{p}{q} + |r|\alpha\right)\right\} \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \left\{j \left(\frac{p}{q} + |r|\alpha\right)\right\} - s_3\left(s\alpha, \frac{p}{q} + |r|\alpha\right) \\
&= \frac{1}{2} - \left(\frac{1}{4} + \frac{d^2}{12s|r|q^2}\right) \\
&= \frac{1}{4} + \frac{d^2}{12srq^2}.
\end{aligned}$$



# 3. Counting function of eigenvalues at microscopic scale

This chapter is based on the article *On a limiting point process related to modified permutation matrices*, not submitted yet.

## Contents

3.1	Introduction . . . . .	77
3.1.1	Spectrum of random permutation matrices . . . . .	77
3.1.2	Main results and outline of the chapter . . . . .	82
3.2	Two natural limiting counting functions . . . . .	84
3.3	Limiting point process related to permutation matrices with modification . . . . .	87
3.4	Continuous analog of the Feller coupling . . . . .	93
3.5	Limiting point process related to permutation matrices . . . . .	98
3.5.1	Approximation of $S$ by $T$ . . . . .	99
3.5.2	Proof of Theorem 3.1.7 (i) . . . . .	100
3.5.3	Proof of Theorem 3.1.7 (ii) . . . . .	103
3.6	Translation of the limiting point process related to permutation matrices	106

## 3.1 Introduction

### 3.1.1 Spectrum of random permutation matrices

Looking at the counting function of eigenvalues of a random permutation matrix, Wieand [87] establishes that the fluctuation of the number of eigenvalues on a fixed arc of the unit circle is asymptotically Gaussian when the size of the matrix goes to infinity, and gives asymptotic expressions of the expectation and variance.

In [89], Wieand tackles more general ensembles of matrices involving random permutations, and shows that her normality result on the fluctuation of the number of eigenvalues holds for these models, with similar behaviors of the expectations and variances.

In these results, Wieand considers uniformly distributed permutations on the symmetric group  $\mathfrak{S}_n$ . Other measures can be relevant to work with. For instance, the family of Ewens measures are of great interest in population genetics, and have many nice properties which make the study of random permutations simple (some of these properties

will be highlighted in the present chapter). Arratia, Barbour and Tavaré [4] give and show many results on Ewens measures. Formally, these measures can be defined in the following way:

Let  $\theta > 0$  and  $n \geq 1$ . A random permutation  $\sigma_n$  of  $\mathfrak{S}_n$  follows the Ewens( $\theta$ ) distribution if

$$\forall \pi \in \mathfrak{S}_n, \mathbb{P}(\sigma_n = \pi) = \mathbb{P}_\theta^{(n)}(\pi) = \frac{\theta^{K(\pi)}}{\theta(\theta+1) \cdots (\theta+n-1)}$$

where  $K(\pi)$  denotes the total number of cycles of  $\pi$  once decomposed into disjoint cycles. The case  $\theta = 1$  corresponds to the uniform measure.

In this chapter we deal with this family of measures on the sets of permutation matrices (we identify the set of the  $n$ -by- $n$  permutation matrices with  $\mathfrak{S}_n$ ). We also consider modifications of these matrices, where the entries equal to one are replaced by complex numbers of modulus one. These modified permutation matrices can be seen as elements of the wreath product  $S^1 \wr \mathfrak{S}_n$ , and for the non-zeros entries we take i.i.d random variables uniformly distributed on the unit circle. One main motivation of taking such a law is to bring closer the analogy with the Circular Unitary Ensemble (the  $n$ -by- $n$  modified permutation matrices form an infinite subgroup of the set of  $n$ -by- $n$  unitary matrices).

A remarkable property that we would like to point out in this work is the invariance of the behavior of the counting function of eigenvalues by change of scale. Indeed, we observe that the leading coefficients in the asymptotic variances are typically the same through the two following approaches:

- Count the eigenvalues in macroscopic or mesoscopic arcs of the unit circle and then let the size of the matrix go to infinity.
- Start from the limiting point process of the microscopic landscape of eigenangles, then count the points in any interval and let the length of this interval tend to infinity.

In order to precise this phenomenon, let us recall a few results which will be helpful for comparison purposes. We use the following notations:

Let  $(M_n)_{n \geq 1}$  be a sequence of random permutation matrices following the Ewens( $\theta$ ) distribution, and let  $(\tilde{M}_n)_{n \geq 1}$  be the sequence of matrices  $M_n$  where the entries equal to one are replaced by i.i.d random variables uniformly distributed on the unit circle. For  $n \geq 1$ , define  $X_n^I$  and  $\tilde{X}_n^I$  as the respective numbers of eigenvalues of  $M_n$  and  $\tilde{M}_n$  which lie in the arc  $I := (e^{2i\pi\alpha}, e^{2i\pi\beta}]$  of the unit circle, for some  $\alpha, \beta$  such that  $0 \leq \alpha < 1$  and  $\alpha < \beta \leq \alpha + 1$ .

For all real numbers  $\alpha$  and  $\beta$ , we set:

$$c_1 = c_1(\alpha, \beta) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n (\{j\beta\} - \{j\alpha\}), \quad (3.1)$$

$$c_2 = c_2(\alpha, \beta) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n (\{j\beta\} - \{j\alpha\})^2, \quad (3.2)$$



$$\ell = \ell(\beta - \alpha) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \{j(\beta - \alpha)\} (1 - \{j(\beta - \alpha)\}), \quad (3.3)$$

where  $\{x\}$  denotes the fractional part of  $x$ . These limits exist and are finite (see *e.g.* Lemma 2.3.8 for a proof).

### Macroscopic scale

The following result has been first established by Wieand [87] [89] in the particular case  $\theta = 1$ , then by Ben Arous and Dang [13] for permutation matrices in the general case  $\theta > 0$ , and can be deduced under stronger assumptions from a result of Dang and Zeindler [34] on the logarithm of the characteristic polynomial of permutation matrices.

**Proposition 3.1.1.** *Let  $0 \leq \alpha < 1$  and  $\alpha < \beta \leq \alpha + 1$ . As  $n \rightarrow \infty$ ,*

$$\begin{aligned} \mathbb{E}(X_n^I) &= n(\beta - \alpha) - \theta c_1 \log n + o(\log n) \\ \text{Var}(X_n^I) &= \theta c_2 \log n + o(\log n) \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}(\tilde{X}_n^I) &= n(\beta - \alpha) \\ \text{Var}(\tilde{X}_n^I) &= \theta \ell \log n + o(\log n). \end{aligned}$$

See Proposition 2.3.4 and Proposition 2.3.7 for a proof of the two last equalities.

### Mesoscopic scale

In Chapter 2, we established the following result:

**Proposition 3.1.2.** *Assume  $I$  to be depending on  $n$ , of the form  $I = I_n := (e^{2i\pi\alpha}, e^{2i\pi(\alpha+\delta_n)})$ , where  $\alpha \in [0, 1)$  and  $(\delta_n)$  is a sequence of positive real numbers satisfying*

$$\begin{cases} \delta_n \xrightarrow[n \rightarrow \infty]{} 0 \\ n\delta_n \xrightarrow[n \rightarrow \infty]{} +\infty. \end{cases}$$

Then, as  $n \rightarrow +\infty$ ,

$$\begin{aligned} \mathbb{E}(X_n^I) &= n\delta_n - \theta c_1 \log(n\delta_n) + o(\log(n\delta_n)) \\ \text{Var}(X_n^I) &= \theta c_2 \log(n\delta_n) + o(\log(n\delta_n)) \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}(\tilde{X}_n^I) &= n\delta_n \\ \text{Var}(\tilde{X}_n^I) &= \theta \ell \log(n\delta_n) + o(\log(n\delta_n)), \end{aligned}$$

with, denoting by  $\kappa$  any arbitrary irrational number,  $c_1 = c_1(\alpha, \kappa)$ ,  $c_2 = c_2(\alpha, \kappa)$ , and  $\ell = \ell(\kappa) = \frac{1}{6}$ .

In fact, the asymptotic of  $\mathbb{E}(X_n^I)$  is not computed in Chapter 2 but can be deduced by the same method as for  $\text{Var}(X_n^I)$ .

Moreover, in both macroscopic and mesoscopic scales, the fluctuations of  $X_n^I$  and  $\tilde{X}_n^I$  are asymptotically Gaussian (Theorems 2.1.2 and 2.1.3).

In this chapter we focus on the microscopic landscape of eigenvalues.

### Microscopic scale

A **virtual permutation** is defined as a sequence of permutation  $\sigma = (\sigma_n)_{n \geq 1}$  where for all  $n$ ,  $\sigma_n \in \mathfrak{S}_n$  and  $\sigma_n$  can be obtained from  $\sigma_{n+1}$  by simply removing the element  $n+1$  in the cycle-decomposition of  $\sigma_{n+1}$ . A remarkable property of the Ewens measures is that if  $\sigma_{n+1}$  follows the  $\text{Ewens}(\theta)$  distribution on  $\mathfrak{S}_{n+1}$ , then  $\sigma_n$  follows the  $\text{Ewens}(\theta)$  distribution on  $\mathfrak{S}_n$ , for every  $\theta > 0$ . Consequently the Ewens measures naturally extend to the space of virtual permutations  $\mathfrak{S}$ .

Let  $\theta > 0$  and let  $\sigma = (\sigma_n)_{n \geq 1}$  be a random virtual permutation following the  $\text{Ewens}(\theta)$  distribution. For  $n \geq 1$ , let  $\ell_{n,j}$  be the length of the  $j$ -th cycle of  $\sigma_n$  in order of appearance (that is to say, in the increasing order of their smallest elements). We complete the sequence  $(\ell_{n,j})_{j \geq 1}$  by zeros. A result of Tsilevitch in [85] states that for all  $j \geq 1$ , as  $n \rightarrow \infty$ ,

$$y_j^{(n)} := \frac{\ell_{n,j}}{n} \xrightarrow{\text{a.s.}} y_j, \quad (3.4)$$

where  $(y_1, y_2, \dots)$  is a random vector following the  $\text{GEM}(\theta)$  distribution. The rearrangement in decreasing order of the coordinates of a  $\text{GEM}(\theta)$  vector follows the Poisson-Dirichlet distribution of parameter  $\theta$  ( $\text{PD}(\theta)$ ), and conversely a size-biased permutation of a  $\text{PD}(\theta)$  vector has  $\text{GEM}(\theta)$  distribution.

For all  $j \geq 1$ ,  $y_j$  has the same law as a product of independent Beta random variables (in the literature this representation of the  $\text{GEM}(\theta)$  distribution is called **stick breaking process**, or **residual allocation model**, see *e.g.* [59] and [75]), and a direct calculation shows that there exist  $r \in (0, 1)$  depending on  $\theta$  and independent on  $j$ , such that

$$\mathbb{E}(y_j) \leq r^j. \quad (3.5)$$

Now, a basic property on permutation matrices is that their eigenvalues are fully determined by the cycle-structure of their associated permutation. More precisely, each  $j$ -cycle of any arbitrary given permutation (once decomposed into disjoint cycles) corresponds to a set of eigenvalues equal to the set of  $j$ -th roots of unity. This supplies us the equalities in distribution

$$X_n^I = \sum_{j=1}^n \mathbb{1}_{\ell_{n,j} > 0} \sum_{w^{\ell_{n,j}} = 1} \mathbb{1}_{w \in I} \quad (3.6)$$

and

$$\tilde{X}_n^I = \sum_{j=1}^n \mathbb{1}_{\ell_{n,j} > 0} \sum_{w^{\ell_{n,j}} = e^{2i\pi\Phi_{n,j}}} \mathbb{1}_{w \in I}, \quad (3.7)$$

where the  $\Phi_{n,j}$  are i.i.d random variables uniformly distributed on  $[0, 1)$ , independent of the  $\ell_{n,j}$ .

Following the same approach as Najnudel and Nikeghbali in [68], since all the eigenvalues of (modified) permutation matrices are on the unit circle, it can be more practical to consider the eigenangles. The corresponding random measures  $\tau(M_n)$  and  $\tau(\widetilde{M}_n)$  can be written as

$$\tau(M_n) = \sum_{j=1}^{\infty} \mathbf{1}_{\ell_{n,j}>0} \sum_{x \equiv 0 \pmod{2\pi/\ell_{n,j}}} \delta_x \quad (3.8)$$

and

$$\tau(\widetilde{M}_n) = \sum_{j=1}^{\infty} \mathbf{1}_{\ell_{n,j}>0} \sum_{x \equiv 2\pi\Phi_{n,j} \pmod{2\pi/\ell_{n,j}}} \delta_x. \quad (3.9)$$

In particular, this immediately implies that  $\tau(M_n)([0, 2\pi)) = \tau(\widetilde{M}_n)([0, 2\pi)) = n$ , in other words, the average spacing of two consecutive points of their respective corresponding point processes is equal to  $2\pi/n$ . Thus, if we want to have a convergence of these measures for  $n$  going to infinity, we need to rescale them in order to have a constant average spacing, say, one. That is why we introduce the rescaled measures  $\tau_n$  and  $\widetilde{\tau}_n$ , defined as the respective images of  $\tau(M_n)$  and  $\tau(\widetilde{M}_n)$  by multiplication by  $n/2\pi$ . One checks that

$$\tau_n = \sum_{j=1}^{+\infty} \mathbf{1}_{y_j^{(n)}>0} \sum_{k \in \mathbb{Z}} \delta_{\frac{k}{y_j^{(n)}}} \quad (3.10)$$

and

$$\widetilde{\tau}_n = \sum_{j=1}^{+\infty} \mathbf{1}_{y_j^{(n)}>0} \sum_{k \in \mathbb{Z}} \delta_{\frac{k+\Phi_{n,j}}{y_j^{(n)}}}. \quad (3.11)$$

Define also the random measures

$$\tau_{\infty} := \sum_{j=1}^{+\infty} \sum_{k \in \mathbb{Z} \setminus \{0\}} \delta_{\frac{k}{y_j}} \quad (3.12)$$

and

$$\widetilde{\tau}_{\infty} := \sum_{j=1}^{+\infty} \sum_{k \in \mathbb{Z}} \delta_{\frac{k+\Phi_j}{y_j}} \quad (3.13)$$

where the  $y_j$  are given by (3.4), and the  $\Phi_j$  are i.i.d random variables uniformly distributed on  $[0, 1)$ , independent of the  $y_j$ .

**Proposition 3.1.3** (Najnudel and Nikeghbali 2010 [68]). *For all continuous functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  with compact support included in  $(0, +\infty)$ ,*

$$\langle \tau_n, f \rangle \xrightarrow[n \rightarrow +\infty]{a.s.} \langle \tau_{\infty}, f \rangle$$

*under the coupling of virtual permutations, and*

$$\langle \widetilde{\tau}_n, f \rangle \xrightarrow[n \rightarrow +\infty]{d} \langle \widetilde{\tau}_{\infty}, f \rangle.$$

In [68] Najnudel and Nikeghbali tackle more general modifications of permutation matrices where the non-zero entries are  $\mathbb{C}$ -valued (not necessarily of modulus one, so that the matrices are no longer unitary). For the wreath product  $S^1 \wr \mathfrak{S}_n$  they also consider more general distributions on  $S^1$  (not necessarily the uniform distribution) and provide analog results on their limiting point processes of eigenvalues.

In the present chapter we will restrain ourselves to the study of the limiting point processes related to  $(M_n)_{n \geq 1}$  and  $(\tilde{M}_n)_{n \geq 1}$ , though the techniques are expected to extend to other ensembles of matrices involving permutations under Ewens measures.

### 3.1.2 Main results and outline of the chapter

In the next section we establish that Proposition 3.1.3 also holds for indicator functions of intervals. This gives a natural meaning to the convergence of the counting function of the microscopic eigenangles, to a limiting counting function. More precisely, we have the following result:

**Proposition 3.1.4.** *For all positive real numbers  $\alpha$  and  $\beta$  such that  $\alpha < \beta$ ,*

$$\tau_n((\alpha, \beta]) \xrightarrow[n \rightarrow +\infty]{a.s.} \tau_\infty((\alpha, \beta])$$

*under the coupling of virtual permutations, and*

$$\tilde{\tau}_n((\alpha, \beta]) \xrightarrow[n \rightarrow +\infty]{d} \tilde{\tau}_\infty((\alpha, \beta]).$$

*Remark 3.1.5.* It is easy to notice that the laws of the measures  $\tilde{\tau}_n$  and  $\tilde{\tau}_\infty$  are invariant by translation. Thus the second point of Proposition 3.1.4 is equivalent to say that for all positive real numbers  $A$ ,  $\tilde{\tau}_n((0, A]) \xrightarrow[n \rightarrow +\infty]{d} \tilde{\tau}_\infty((0, A])$ . Moreover, the choice of including or excluding the endpoints of the interval  $(0, A]$  does not have importance for  $\tilde{\tau}_n$  since for all  $x \in \mathbb{R}$ ,  $\tilde{\tau}_n(x) = 0$  almost surely. This is clearly not true for  $\tau_n$ , but it can be proven that for all fixed  $x > 0$ ,  $\tau_n(x) \rightarrow 0 = \tau_\infty(x)$  almost surely as  $n \rightarrow \infty$  under the coupling of virtual permutations. Indeed,  $\tau_n(x) = \sum_{j \geq 1: \ell_{n,j} > 0} \mathbb{1}_{xy_j^{(n)} \in \mathbb{Z}}$ , so if  $0 < x < 1$  we have  $\tau_n(x) = 0$  (since  $y_j^{(n)} \in (0, 1]$  for all  $j$  such that  $\ell_{n,j} > 0$ ) and if  $x \geq 1$  we have for all  $j$ ,

$$\mathbb{1}_{y_j^{(n)} > 0, xy_j^{(n)} \in \mathbb{Z}} \leq \mathbb{1}_{xy_j^{(n)} \geq 1} \leq \mathbb{1}_{\sup_n y_j^{(n)} \geq 1/x} \leq \mathbb{1}_{C\rho^j \geq 1/x}$$

(see Lemma 3.2.1 for the last inequality) which is summable, and then by dominated convergence we get  $\tau_n(x) \rightarrow 0$  a.s., since  $xy_j \notin \mathbb{Z}$  a.s. and then  $\mathbb{1}_{xy_j^{(n)} \in \mathbb{Z}} \rightarrow 0$  a.s. for each  $j \geq 1$ .

More generally, Proposition 3.1.4 extends to finite numbers of intervals, which immediately implies that both convergences hold for finite combinations of indicator functions.

Now, we present our two main results, involving  $\tau_\infty$  and  $\tilde{\tau}_\infty$ :

**Theorem 3.1.6.** *Let  $A > 1$ .*

$$\frac{\tilde{\tau}_\infty([0, A]) - A}{\sqrt{\frac{\theta}{6} \log A}} \xrightarrow[A \rightarrow +\infty]{d} \mathcal{N}(0, 1).$$

**Theorem 3.1.7.** *Let  $a$  and  $b$  be two positive real numbers such that  $a < b$ ,*

(i) *As  $b \rightarrow +\infty$ ,*

$$\mathbb{E}(\tau_\infty((a, a+b])) = b - \frac{\theta}{2} \log b + \mathcal{O}_\theta(1)$$

*and*

$$\text{Var}(\tau_\infty((a, a+b])) = \frac{\theta}{3} \log b + \mathcal{O}_\theta(\sqrt{\log b}).$$

*Moreover,*

$$\frac{\tau_\infty((a, a+b]) - \mathbb{E}(\tau_\infty((a, a+b]))}{\sqrt{\text{Var}(\tau_\infty((a, a+b]))}} \xrightarrow{d} \mathcal{N}(0, 1).$$

(ii) *Let  $\nu$  be a real number greater than 1. As  $a \rightarrow +\infty$  and  $b = (\nu - 1)a$ ,*

$$\mathbb{E}(\tau_\infty((a, \nu a])) = (\nu - 1)a + \mathcal{O}_\theta(1)$$

*and*

$$\begin{aligned} & \text{Var}(\tau_\infty((a, \nu a])) \\ &= \begin{cases} \frac{\theta}{6} (1 - \frac{1}{rs}) \log a + \mathcal{O}_\theta(\sqrt{\log a}) & \text{if } \nu = \frac{r}{s} \text{ with } \gcd(r, s) = 1 \\ \frac{\theta}{6} \log a + \mathcal{O}_\theta(\sqrt{\log a}) & \text{if } \nu \text{ is irrational.} \end{cases} \end{aligned}$$

*Moreover,*

$$\frac{\tau_\infty((a, \nu a]) - \mathbb{E}(\tau_\infty((a, \nu a]))}{\sqrt{\text{Var}(\tau_\infty((a, \nu a]))}} \xrightarrow{d} \mathcal{N}(0, 1).$$

*Remark 3.1.8.* Note that Theorem 3.1.7 can be related to Propositions 3.1.1 and 3.1.2. In fact, the coefficients in the asymptotic expressions of the expectation and of the variance behave similarly, in the following sense:

- Point (i) is linked to the case of a macroscopic arc of the form  $I = (e^{2i\pi\alpha}, e^{2i\pi\beta}]$  with  $\alpha = 0$  and  $\beta$  irrational, and also to the case of a mesoscopic arc with the same  $\alpha = 0$  and replacing  $\beta$  by  $\delta_n$  (where  $\delta_n$  decreases to 0 slower than  $1/n$  as  $n$  goes to  $\infty$ ). Indeed, a direct computation (see [87]) of  $c_1$  and  $c_2$  gives  $c_1 = \frac{1}{2}$  and  $c_2 = \frac{1}{3}$  for this particular case.
- Point (ii) is linked to the case of a macroscopic arc of the form  $I = (e^{2i\pi\alpha}, e^{2i\pi\beta}]$  with  $\alpha$  irrational and  $\beta$  irrational, and also to the case of a mesoscopic arc with  $\alpha$  irrational and  $\beta = \alpha + \delta_n$ . Indeed, a direct computation (see [87] and Chapter 2 Appendix B) of  $c_1$  and  $c_2$  gives  $c_1 = 0$  and

$$c_2 = \begin{cases} \frac{1}{6} (1 - \frac{1}{rs}) & \text{if } \beta = \frac{r}{s}\alpha \text{ with } \gcd(r, s) = 1 \text{ and } \frac{r}{s} > 1 \\ \frac{1}{6} & \text{if } \alpha \text{ and } \beta \text{ are } \mathbb{Z}\text{-linearly independent.} \end{cases}$$

The empirical measures  $\tau_\infty$  and  $\tilde{\tau}_\infty$  are related to each other by the following special link:

**Proposition 3.1.9.** *Let  $f \in \mathcal{C}(\mathbb{R}, \mathbb{C})$  with compact support. Let  $A > 0$ . Then*

$$\langle \tau_\infty \circ T_A, f \rangle \xrightarrow[A \rightarrow +\infty]{d} \langle \tilde{\tau}_\infty, f \rangle,$$

where  $T_A$  is the shift operator defined by  $T_A : x \mapsto x + A$ .

The chapter follows a linear structure: In Section 3.2 we motivate the study of the considered limiting objects and give a proof of Proposition 3.1.4. In Section 3.3 we prove Theorem 3.1.6. In Section 3.4, we introduce a main tool that we use in Section 3.5 for proving Theorem 3.1.7. This tool is an analog of the ubiquitous Feller coupling, and has interest beyond our study. Finally, in Section 3.6 we prove Proposition 3.1.9.

### 3.2 Two natural limiting counting functions. Proof of Proposition 3.1.4

We begin with the following lemma:

**Lemma 3.2.1.** *There exist  $\rho \in (0, 1)$  depending on  $\theta$ , and a random number  $C > 0$  such that a.s., for all  $j \geq 1$ ,*

$$s_j := \sup_{m \geq 1} y_j^{(m)} \leq C \rho^j. \quad (3.14)$$

*Proof.* First, it can be checked that for all  $j$ , the sequence  $\left(\frac{\ell_{N,j}}{N+\theta}\right)_{N \geq 1}$  is a submartingale with respect to the filtration  $(\mathcal{F}_N)$  (see e.g. [85] for a proof), where  $\mathcal{F}_N$  is the  $\sigma$ -algebra generated by  $(\ell_{p,q}, 1 \leq p \leq N, q \leq p)$ . Moreover, as this submartingale is positive and bounded in  $L^2$  (clear since the terms are bounded by 1), then it follows from Doob's inequality

$$\mathbb{E} \left( \left( \sup_{N \geq 1} \frac{\ell_{N,j}}{N+\theta} \right)^2 \right) \leq 4 \sup_{N \geq 1} \mathbb{E} \left( \left( \frac{\ell_{N,j}}{N+\theta} \right)^2 \right)$$

and then, since  $\frac{\ell_{N,j}}{N+\theta}$  is lower than 1,

$$\begin{aligned} \mathbb{E}(s_j^2) &\leq 4(1+\theta)^2 \sup_{N \geq 1} \mathbb{E} \left( \frac{\ell_{N,j}}{N+\theta} \right) \\ &= 4(1+\theta)^2 \lim_{N \rightarrow +\infty} \mathbb{E} \left( \frac{\ell_{N,j}}{N+\theta} \right) \\ &\leq 4(1+\theta)^2 \lim_{N \rightarrow +\infty} \mathbb{E}(y_j^{(N)}) \\ &= 4(1+\theta)^2 \mathbb{E} \left( \lim_{N \rightarrow +\infty} y_j^{(N)} \right) \\ &= 4(1+\theta)^2 \mathbb{E}(y_j) \\ &\leq 4(1+\theta)^2 r^j. \end{aligned}$$

where we use the submartingale property for the first equality, the dominated convergence theorem for the second and third equalities, and (3.5) for the last inequality.

Moreover, using Cauchy-Schwarz inequality we deduce  $\mathbb{E}(s_j) \leq 2(1 + \theta)r^{j/2}$ , and finally  $\rho := \frac{1+r^{1/2}}{2} \in (0, 1)$  gives

$$\mathbb{P}(s_j > \rho^j) \leq \frac{1}{\rho^j} \mathbb{E}(s_j) \leq 2(1 + \theta) \left( \frac{r^{1/2}}{\rho} \right)^j$$

which is summable in  $j$ , therefore Borel-Cantelli lemma applies.  $\square$

Let  $\alpha$  and  $\beta$  two real numbers such that  $0 \leq \alpha < 1$  and  $\alpha < \beta \leq \alpha + 1$ . For all  $n$ , the random numbers  $X_n^I$  and  $\tilde{X}_n^I$  of eigenvalues of  $M_n$  and  $\tilde{M}_n$  lying in the arc  $I(e^{2i\pi\alpha}, e^{2i\pi\beta}]$  are given by the following expressions (see [87]):

$$\begin{aligned} X_n^I &= \sum_{j=1}^n \mathbb{1}_{\ell_{n,j} > 0} (\lfloor \ell_{n,j} \beta \rfloor - \lfloor \ell_{n,j} \alpha \rfloor) \\ &= n(\beta - \alpha) - \sum_{j=1}^n \mathbb{1}_{\ell_{n,j} > 0} (\{\ell_{n,j} \beta\} - \{\ell_{n,j} \alpha\}) \end{aligned}$$

and

$$\begin{aligned} \tilde{X}_n^I &= \sum_{j=1}^n \mathbb{1}_{\ell_{n,j} > 0} (\lfloor \ell_{n,j} \beta - \Phi_{n,j} \rfloor - \lfloor \ell_{n,j} \alpha - \Phi_{n,j} \rfloor) \\ &= n(\beta - \alpha) - \sum_{j=1}^n \mathbb{1}_{\ell_{n,j} > 0} (\{\ell_{n,j} \beta\} - \{\ell_{n,j} \alpha\} - \mathbb{1}_{\Phi_{n,j} \leq \{\ell_{n,j} \beta\}} + \mathbb{1}_{\Phi_{n,j} \leq \{\ell_{n,j} \alpha\}}) \end{aligned}$$

where  $(\Phi_{n,j})_{n,j \geq 1}$  is an array of i.i.d random variables uniformly distributed on  $[0, 1)$ , independent of  $(\sigma_n)_{n \geq 1}$ .

If we replace  $\alpha$  and  $\beta$  respectively by  $\alpha/n$  and  $\beta/n$ , we get

$$X_n^{(e^{2i\pi\alpha/n}, e^{2i\pi\beta/n})} = \beta - \alpha - \sum_{j=1}^n \mathbb{1}_{y_j^{(n)} > 0} (\{y_j^{(n)} \beta\} - \{y_j^{(n)} \alpha\})$$

and

$$\begin{aligned} \tilde{X}_n^{(e^{2i\pi\alpha/n}, e^{2i\pi\beta/n})} &= \beta - \alpha - \sum_{j=1}^n \mathbb{1}_{y_j^{(n)} > 0} \left( \{y_j^{(n)} \beta\} - \{y_j^{(n)} \alpha\} - \mathbb{1}_{\Phi_{n,j} \leq \{y_j^{(n)} \beta\}} + \mathbb{1}_{\Phi_{n,j} \leq \{y_j^{(n)} \alpha\}} \right). \end{aligned}$$

From this it seems reasonable to consider the version  $n = \infty$  of these quantities, in order to count the points of the limiting point process obtained as the limit of the sequence of eigenangles multiplied by  $n/2\pi$  (microscopic scale). The following proposition gives a meaning to the convergence.

**Proposition 3.2.2.** *We have the following convergences:*

$$X_n^{(e^{2i\pi\alpha/n}, e^{2i\pi\beta/n})} \xrightarrow[n \rightarrow \infty]{a.s.} \beta - \alpha - \sum_{j=1}^{+\infty} (\{y_j\beta\} - \{y_j\alpha\}) \quad (3.15)$$

under the coupling of virtual permutations, and

$$\tilde{X}_n^{(e^{2i\pi\alpha/n}, e^{2i\pi\beta/n})} \xrightarrow[n \rightarrow \infty]{d} \beta - \alpha - \sum_{j=1}^{+\infty} \left( \{y_j\beta\} - \{y_j\alpha\} - \mathbb{1}_{\Phi_j \leq \{y_j\beta\}} + \mathbb{1}_{\Phi_j \leq \{y_j\alpha\}} \right) \quad (3.16)$$

where the  $\Phi_j$  are i.i.d. random variables uniformly distributed on  $[0, 1)$ , independent of the  $y_i$ .

*Remark 3.2.3.* Note that this proposition is a reformulation of Proposition 3.1.4.

*Proof.* First, we know that a.s., for all  $j$ ,  $y_j > 0$ , hence  $\mathbb{1}_{y_j^{(n)} > 0} \xrightarrow[n \rightarrow \infty]{} 1$ .

Let  $x > 0$ . We are going to show that a.s.,  $\sum_{j=1}^{+\infty} \{y_j^{(n)}x\} \xrightarrow[n \rightarrow \infty]{} \sum_{j=1}^{+\infty} \{y_jx\}$ . By Lemma 3.2.1, almost surely there exists  $\rho \in (0, 1)$  and a random number  $C > 0$  such that for all  $j$  and  $n$ ,  $y_j^{(n)} \leq C\rho^j$ , then

$$\exists j_0 \in \mathbb{N}^*, \quad \forall j > j_0, \quad \forall n \geq 1, \quad y_j^{(n)}x \leq \frac{1}{2}.$$

Fix  $j_0$ . Letting  $n$  tend to infinity, as  $y_j^{(n)} \xrightarrow[n \rightarrow \infty]{} y_j$  a.s., we have for all  $j > j_0$ ,  $y_jx \leq \frac{1}{2}$  and then

$$|\{y_j^{(n)}x\} - \{y_jx\}| = |x(y_j^{(n)} - y_j)| \xrightarrow[n \rightarrow \infty]{} 0.$$

Moreover, obviously for all  $j$  and  $n$ ,

$$\{y_j^{(n)}x\} \leq y_j^{(n)}x \leq Cx\rho^j$$

which is summable in  $j$ . Hence, by dominated convergence it follows

$$\sum_{j=j_0+1}^{+\infty} \{y_j^{(n)}x\} \xrightarrow[n \rightarrow \infty]{} \sum_{j=j_0+1}^{+\infty} \{y_jx\} \quad (3.17)$$

almost surely.

For  $j \leq j_0$  the idea is to take  $n$  large enough such that the only  $y_jx$  that could pose a challenge are integers (discontinuities of the fractional part function). Let  $\varepsilon > 0$ . There exists  $N \in \mathbb{N}^*$  such that for all  $n \geq N$ , for all  $j \leq j_0$ ,

$$|\{y_j^{(n)}x\} - \{y_jx\}| \leq \begin{cases} \frac{\varepsilon}{j_0} & \text{if } y_jx \notin \mathbb{N} \\ 1 & \text{if } y_jx \in \mathbb{N} \end{cases},$$

and then

$$\sum_{j=1}^{j_0} |\{y_j^{(n)}x\} - \{y_jx\}| \leq \varepsilon + \sum_{j=1}^{j_0} \mathbb{1}_{y_jx \in \mathbb{N}}.$$



In addition  $\sum_{j \leq j_0} \mathbb{1}_{y_j x \in \mathbb{N}} = 0$  a.s. since it is a finite sum of indicators of negligible events. From (3.17) we deduce that a.s.,

$$\sum_{j=1}^{+\infty} \{y_j^{(n)} x\} \xrightarrow{n \rightarrow \infty} \sum_{j=1}^{+\infty} \{y_j x\}.$$

It just remains to prove the convergence in distribution of  $Q_n := \sum_{j=1}^{+\infty} \left( \mathbb{1}_{\Phi_{n,j} \leq \{y_j^{(n)} \beta\}} - \mathbb{1}_{\Phi_{n,j} \leq \{y_j^{(n)} \alpha\}} \right)$  to  $Q := \sum_{j=1}^{+\infty} \left( \mathbb{1}_{\Phi_j \leq \{y_j \beta\}} - \mathbb{1}_{\Phi_j \leq \{y_j \alpha\}} \right)$ .

Let  $t \in \mathbb{R}$ . Denoting  $\omega_{j,n} := \{y_j^{(n)} \beta\} - \{y_j^{(n)} \alpha\}$ , we have:

$$\begin{aligned} \mathbb{E} \left[ e^{itQ_n} | (y_j^{(m)})_{j,m \geq 1} \right] &= \prod_{j=1}^{+\infty} (e^{it\omega_{j,n}} \mathbb{1}_{\omega_{j,n} > 0} + e^{-it} (-\omega_{j,n}) \mathbb{1}_{\omega_{j,n} < 0} + 1 \times (1 - |\omega_{j,n}|)) \\ &= \prod_{j=1}^{+\infty} (1 + (e^{it} - 1)\omega_{j,n} \mathbb{1}_{\omega_{j,n} > 0} - (e^{-it} - 1)\omega_{j,n} \mathbb{1}_{\omega_{j,n} < 0}). \end{aligned}$$

Taking the logarithm for  $j$  large enough, and noting that a.s. there is no  $j$  such that  $y_j \alpha$  or  $y_j \beta$  is integer ( $\alpha, \beta > 0$ ), the dominated convergence theorem ensures that

$$\mathbb{E} \left[ e^{itQ_n} | (y_j^{(m)})_{j,m \geq 1} \right] \xrightarrow{n \rightarrow \infty} \mathbb{E} \left[ e^{itQ} | (y_j)_{j \geq 1} \right]$$

for almost every realization of  $(y_j)_{j \geq 1}$ . Applying once again the dominated convergence theorem, we get

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[ \mathbb{E} \left[ e^{itQ_n} | (y_j^{(m)})_{j,m \geq 1} \right] \right] = \mathbb{E} \left[ \lim_{n \rightarrow \infty} \mathbb{E} \left[ e^{itQ_n} | (y_j)_{j \geq 1} \right] \right] = \mathbb{E} \left[ e^{itQ} \right].$$

□

### 3.3 Limiting point process related to permutation matrices with modification. Proof of Theorem 3.1.6

For  $A > 0$ , define

$$\tilde{X}(A) = A + \sum_{j=1}^{+\infty} \left( \mathbb{1}_{\Phi_j \leq \{Ay_j\}} - \{Ay_j\} \right).$$

According to the previous section, this random variable counts the number of points in  $[0, A]$  of the limiting point process of normalized eigenangles of  $\tilde{M}_n$  when  $n$  goes to infinity, i.e we have  $\tilde{X}(A) = \tilde{\tau}_\infty([0, A])$ . Then, proving Theorem 3.1.6 amounts to show

$$\frac{\tilde{X}(A) - A}{\sqrt{\frac{\theta}{6} \log(A)}} \xrightarrow[A \rightarrow \infty]{d} \mathcal{N}(0, 1). \quad (3.18)$$

### 3. COUNTING FUNCTION OF EIGENVALUES AT MICROSCOPIC SCALE

---

Let  $A > 1$ . We first notice that we can write

$$\tilde{X}(A) - A = \sum_{j=1}^{+\infty} B(p_j)$$

where the  $B(p_j)$  are centred Bernoulli random variables of random parameters  $p_j := \{Ay_j\}$ , which are independent conditionally on the  $y_j$ .

Let  $\lambda_0 \in \mathbb{R}$  and denote  $\lambda := \frac{\lambda_0}{\sqrt{\frac{\theta}{6} \log(A)}}$ .

$$\begin{aligned} \mathbb{E} \left[ e^{i\lambda \sum_{j=1}^{+\infty} B(p_j)} \mid (y_m)_{m \geq 1} \right] &= \prod_{j=1}^{+\infty} \mathbb{E} \left[ e^{i\lambda B(p_j)} \mid (y_m)_{m \geq 1} \right] \\ &= \prod_{j=1}^{+\infty} \left( 1 + p_j \left( e^{i\lambda(1-p_j)} - 1 \right) + (1-p_j) \left( e^{-i\lambda p_j} - 1 \right) \right) \\ &\underset{A \rightarrow \infty}{=} \prod_{j=1}^{+\infty} \left( 1 - \frac{\lambda^2}{2} p_j(1-p_j)(1 + \mathcal{O}(\lambda)) \right). \end{aligned}$$

Moreover, since the sequence  $(p_j(1-p_j))_{j \geq 1}$  is bounded (uniformly in  $A$ ) and using the fact that for all complex numbers  $z$  sufficiently close to zero we have  $1 + z = \exp(z + \mathcal{O}(z^2))$ , it follows that for all  $A$  large enough,

$$\mathbb{E} \left[ e^{i\lambda \sum_{j=1}^{+\infty} B(p_j)} \right] = \mathbb{E} \left( \exp \left( -\frac{\lambda^2}{2} (1 + \mathcal{O}(\lambda)) \sum_{j=1}^{+\infty} p_j(1-p_j) \right) \right).$$

Thus we want to show that

$$\mathbb{E} \left( \exp \left( -\frac{\lambda^2}{2} (1 + \mathcal{O}(\lambda)) \sum_{j=1}^{+\infty} p_j(1-p_j) \right) \right) \xrightarrow{A \rightarrow \infty} e^{-\frac{\lambda_0^2}{2}}.$$

For this purpose, it suffices to show that the random variable  $Z_A := \frac{1}{\log A} \sum_{j=1}^{+\infty} p_j(1-p_j)$  converges in probability to  $\frac{\theta}{6}$  when  $A$  goes to  $+\infty$ . Indeed, if we show this, then  $Z_A(1 + \mathcal{O}(\lambda))$  will clearly converge in probability to  $\frac{\theta}{6}$  and it will just remain to apply the definition of the convergence in distribution of  $Z_A(1 + \mathcal{O}(\lambda))$  (which is positive for all  $A$  large enough) to the bounded continuous function  $f : x \mapsto \exp \left( -\frac{3\lambda_0^2}{\theta} x \right)$  on  $[0, +\infty)$ . Let  $\varepsilon > 0$ . We cut the sum in  $Z_A$  into three parts:  $j > (1 + \varepsilon)\theta \log A$ ,  $(1 - \varepsilon)\theta \log A < j \leq (1 + \varepsilon)\theta \log A$  and  $j \leq (1 - \varepsilon)\theta \log A$ .

In the first regime, we have, noticing that for all integers  $k \geq 1$ ,  $\sum_{j=k+1}^{+\infty} y_j \stackrel{d}{=} \prod_{j=1}^k U_j$  where the random variables  $U_j$  are independent and follow Beta distribution of parameters  $\theta$

and 1,

$$\begin{aligned} \mathbb{P} \left( \sum_{j > (1+\varepsilon)\theta \log A} p_j(1-p_j) \geq 1 \right) &\leq \mathbb{P} \left( \sum_{j > (1+\varepsilon)\theta \log A} Ay_j \geq 1 \right) \\ &= \mathbb{P} \left( \prod_{j \leq (1+\varepsilon)\theta \log A} U_j \geq \frac{1}{A} \right) \\ &= \mathbb{P} \left( \frac{1}{(1+\varepsilon)\theta \log A} \sum_{j \leq (1+\varepsilon)\theta \log A} \log U_j \geq -\frac{1}{(1+\varepsilon)\theta} \right). \end{aligned}$$

As  $\mathbb{E}(\log U_1) = \int_0^1 \log(x) \theta x^{\theta-1} dx = \frac{-1}{\theta}$  and  $\frac{-1}{(1+\varepsilon)\theta} > \frac{-1}{\theta}$ , then the weak law of large numbers yields

$$\mathbb{P} \left( \sum_{j > (1+\varepsilon)\theta \log A} p_j(1-p_j) \geq 1 \right) \xrightarrow{A \rightarrow \infty} 0,$$

and then

$$\frac{1}{\log A} \sum_{j > (1+\varepsilon)\theta \log A} p_j(1-p_j) \xrightarrow{\mathbb{P}} 0. \quad (3.19)$$

For the  $j$  satisfying  $(1-\varepsilon)\theta \log A < j \leq (1+\varepsilon)\theta \log A$ ,

$$\begin{aligned} \frac{1}{\log A} \sum_{(1-\varepsilon)\theta \log A < j \leq (1+\varepsilon)\theta \log A} p_j(1-p_j) &\leq \frac{1}{\log A} \sum_{(1-\varepsilon)\theta \log A < j \leq (1+\varepsilon)\theta \log A} 1 \\ &< 2\theta\varepsilon + \frac{1}{\log A}. \end{aligned} \quad (3.20)$$

Finally, for  $j \leq (1-\varepsilon)\theta \log A$ , let us show that the sum converges in probability to  $\frac{\theta}{6}(1-\varepsilon)$ . To this end, it is enough to show that its two first moments respectively converge to  $\frac{\theta}{6}(1-\varepsilon)$  and  $\left(\frac{\theta}{6}(1-\varepsilon)\right)^2$ .

Recall that for all  $j$ ,  $p_j = \{Ay_j\}$ , so the computation of the moments is not obvious. Note that  $p_j(1-p_j) = \frac{1}{6} - B_2(p_j)$ , where  $B_2$  is the second Bernoulli polynomial ( $B_2(x) = x^2 - x + \frac{1}{6}$ ), which gives a simple expression of its Fourier series. More precisely, for all  $x \in \mathbb{R}$  we have the following expansion in Fourier series:

$$\{x\}(1-\{x\}) = \frac{1}{6} - \frac{1}{2\pi^2} \sum_{k \neq 0} \frac{e^{2i\pi kx}}{k^2}.$$

Hence,

$$\frac{1}{\log A} \sum_{j \leq (1-\varepsilon)\theta \log A} p_j(1-p_j) = \frac{\lfloor (1-\varepsilon)\theta \log A \rfloor}{6 \log A} - \frac{1}{2\pi^2 \log A} \sum_{j \leq (1-\varepsilon)\theta \log A} \sum_{k \neq 0} \frac{e^{2i\pi kAy_j}}{k^2}.$$

For  $k \neq 0$ ,

$$\begin{aligned}\mathbb{E}\left(e^{2i\pi k A y_j}\right) &= \mathbb{E}\left[\mathbb{E}\left[e^{2i\pi k A U_1 \dots U_{j-1}(1-U_j)} \mid (U_m)_{m \leq j-1}\right]\right] \\ &= \mathbb{E}\left[\int_0^1 e^{2i\pi k A U_1 \dots U_{j-1}(1-x)} \theta x^{\theta-1} dx\right]\end{aligned}$$

Let  $\alpha \in (1 - \varepsilon, 1)$  and  $\eta \in (0, 1)$  that we will precise at the end of the proof.

We write

$$\begin{aligned}\int_0^1 e^{2i\pi k A U_1 \dots U_{j-1}(1-x)} \theta x^{\theta-1} dx &= \int_0^1 e^{2i\pi k A U_1 \dots U_{j-1}(1-x)} \theta x^{\theta-1} dx \mathbb{1}_{U_1 \dots U_{j-1} \leq A^{-\alpha}} \\ &\quad + \int_0^\eta e^{2i\pi k A U_1 \dots U_{j-1}(1-x)} \theta x^{\theta-1} dx \mathbb{1}_{U_1 \dots U_{j-1} > A^{-\alpha}} \\ &\quad + \int_\eta^1 e^{2i\pi k A U_1 \dots U_{j-1}(1-x)} \theta x^{\theta-1} dx \mathbb{1}_{U_1 \dots U_{j-1} > A^{-\alpha}}\end{aligned}$$

For the first term on the right-hand side of the equality,

$$\begin{aligned}\mathbb{E}\left[\left|\int_0^1 e^{2i\pi k A U_1 \dots U_{j-1}(1-x)} \theta x^{\theta-1} dx \mathbb{1}_{U_1 \dots U_{j-1} \leq A^{-\alpha}}\right|\right] \\ \leq \int_0^1 \theta x^{\theta-1} dx \mathbb{P}(U_1 \dots U_{j-1} \leq A^{-\alpha}) \\ \leq \mathbb{P}\left(\prod_{m \leq (1-\varepsilon)\theta \log A} U_m \leq A^{-\alpha}\right) \\ = \mathbb{P}\left(\frac{1}{(1-\varepsilon)\theta \log A} \sum_{m \leq (1-\varepsilon)\theta \log A} \log U_m \leq -\frac{\alpha}{(1-\varepsilon)\theta}\right) \\ \xrightarrow[A \rightarrow +\infty]{} 0\end{aligned}$$

by the weak law of large numbers, since  $\frac{-\alpha}{(1-\varepsilon)\theta} < \mathbb{E}(\log U_1) = \frac{-1}{\theta}$ . Note that the convergence is uniform in  $j$  and  $k$ .

For the second term,

$$\left|\int_0^\eta e^{2i\pi k A U_1 \dots U_{j-1}(1-x)} \theta x^{\theta-1} dx \mathbb{1}_{U_1 \dots U_{j-1} > A^{-\alpha}}\right| \leq \int_0^\eta \theta x^{\theta-1} dx = \eta^\theta.$$

For the third term, an integration by parts gives

$$\begin{aligned}\int_\eta^1 e^{2i\pi k A U_1 \dots U_{j-1}(1-x)} \theta x^{\theta-1} dx &= \left[-\frac{e^{2i\pi k A U_1 \dots U_{j-1}(1-x)}}{2i\pi k A U_1 \dots U_{j-1}} \theta x^{\theta-1}\right]_\eta^1 \\ &\quad + \int_\eta^1 \frac{e^{2i\pi k A U_1 \dots U_{j-1}(1-x)}}{2i\pi k A U_1 \dots U_{j-1}} \theta(\theta-1) x^{\theta-2} dx,\end{aligned}$$

so that

$$\begin{aligned} & \left| \int_{\eta}^1 e^{2i\pi k A U_1 \dots U_{j-1}(1-x)} \theta x^{\theta-1} dx \mathbf{1}_{U_1 \dots U_{j-1} > A^{-\alpha}} \right| \\ & \leq \frac{2\theta}{2\pi|k| A U_1 \dots U_{j-1}} (1 + \eta^{\theta-1}) \mathbf{1}_{U_1 \dots U_{j-1} > A^{-\alpha}} \\ & \leq \frac{\theta}{\pi} (1 + \eta^{\theta-1}) A^{\alpha-1}. \end{aligned}$$

It remains to show that we can chose  $\eta$  (depending on  $A$ ) such that  $\max(\eta^{\theta}, \eta^{\theta-1} A^{\alpha-1})$  converges to 0 when  $A$  goes to infinity. If  $\theta \geq 1$  it is clear, for instance we can take  $\eta = A^{-1}$ . If  $\theta < 1$ ,  $\eta = A^{\frac{1-\alpha}{2(\theta-1)}}$  works.

We deduce

$$\mathbb{E} \left( e^{2i\pi k A y_j} \right)_{A \rightarrow +\infty} = o(1)$$

where the  $o(1)$  is independent of  $k$  and  $j$ . Consequently,

$$\begin{aligned} & \mathbb{E} \left[ \frac{1}{\log A} \sum_{j \leq (1-\varepsilon)\theta \log A} p_j(1-p_j) \right] \\ & \stackrel{A \rightarrow +\infty}{=} \left( \frac{\theta}{6}(1-\varepsilon) + o(1) \right) - \left( \frac{1}{2\pi^2 \log A} \sum_{j \leq (1-\varepsilon)\theta \log A} \sum_{k \neq 0} \frac{1}{k^2} \right) o(1) \\ & = \frac{\theta}{6}(1-\varepsilon) + o(1). \end{aligned}$$

Now, let us show that the second moment converges to  $(\frac{\theta}{6}(1-\varepsilon))^2$ . We have

$$\begin{aligned} & \left( \sum_{j \leq (1-\varepsilon)\theta \log A} p_j(1-p_j) \right)^2 \\ & = \frac{[(1-\varepsilon)\theta \log A]^2}{36} - \frac{[(1-\varepsilon)\theta \log A]}{6\pi^2} \sum_{j \leq (1-\varepsilon)\theta \log A} \sum_{k \neq 0} \frac{e^{2i\pi k A y_j}}{k^2} \\ & \quad + \frac{1}{4\pi^4} \sum_{j_1, j_2 \leq (1-\varepsilon)\theta \log A} \sum_{k, l \neq 0} \frac{e^{2i\pi A(ky_{j_1} + ly_{j_2})}}{k^2 l^2}. \end{aligned}$$

Let  $j_1, j_2 \geq 1$  and  $k, l \neq 0$ .

- If  $j_2 > j_1$ , then

$$\begin{aligned} & \mathbb{E} \left( e^{2i\pi A(ky_{j_1} + ly_{j_2})} \right) \\ & = \mathbb{E} \left[ \mathbb{E} \left[ e^{2i\pi A k U_1 \dots U_{j_1-1}(1-U_{j_1})} e^{2i\pi A l U_1 \dots U_{j_2-1}(1-U_{j_2})} \mid (U_m)_{m \leq j_2-1} \right] \right] \\ & = \mathbb{E} \left[ e^{2i\pi A k U_1 \dots U_{j_1-1}(1-U_{j_1})} \int_0^1 e^{2i\pi A l U_1 \dots U_{j_2-1}(1-x)} \theta x^{\theta-1} dx \right] \end{aligned}$$

and

$$\begin{aligned} & \left| e^{2i\pi A k U_1 \dots U_{j_1-1}(1-U_{j_1})} \int_0^1 e^{2i\pi A l U_1 \dots U_{j_2-1}(1-x)} \theta x^{\theta-1} dx \right| \\ &= \left| \int_0^1 e^{2i\pi A l U_1 \dots U_{j_2-1}(1-x)} \theta x^{\theta-1} dx \right| \end{aligned}$$

so, dividing into three pieces as previously we get  $\mathbb{E} \left( e^{2i\pi A(ky_{j_1} + ly_{j_2})} \right) = o(1)$  where the  $o(1)$  is independent of  $k, l, j_1$  and  $j_2$ .

- If  $j_1 = j_2$  and  $k + l \neq 0$ , then

$$\mathbb{E} \left( e^{2i\pi A(ky_{j_1} + ly_{j_2})} \right) = \mathbb{E} \left( e^{2i\pi A(k+l)y_{j_1}} \right) = o(1)$$

as above.

- If  $j_1 = j_2$  and  $k + l = 0$ , then

$$\mathbb{E} \left( e^{2i\pi A(ky_{j_1} + ly_{j_2})} \right) = 1.$$

Thus,

$$\begin{aligned} \mathbb{E} \left( \sum_{j_1, j_2 \leq (1-\varepsilon)\theta \log A} \sum_{k, l \neq 0} \frac{e^{2i\pi A(ky_{j_1} + ly_{j_2})}}{k^2 l^2} \right) &= o((\log A)^2) + \sum_{j_1 \leq (1-\varepsilon)\theta \log A} \sum_{k \neq 0} \frac{1}{k^2 (-k)^2} \\ &= o((\log A)^2), \end{aligned}$$

and it follows

$$\mathbb{E} \left[ \left( \frac{1}{\log A} \sum_{j \leq (1-\varepsilon)\theta \log A} p_j (1 - p_j) \right)^2 \right] \underset{A \rightarrow +\infty}{=} \left( \frac{\theta}{6} (1 - \varepsilon) \right)^2 + o(1).$$

Consequently,

$$\frac{1}{\log A} \sum_{j \leq (1-\varepsilon)\theta \log A} p_j (1 - p_j) \xrightarrow{\mathbb{P}} \frac{\theta}{6} (1 - \varepsilon). \quad (3.21)$$

Let us now finish to prove the convergence in probability of  $Z_A$  to  $\frac{\theta}{6}$ . For the sake of simplicity, denote

$$\begin{cases} Z_{A, >} := \frac{1}{\log A} \sum_{j > (1+\varepsilon)\theta \log A} p_j (1 - p_j) \\ Z_{A, \star} := \frac{1}{\log A} \sum_{(1-\varepsilon)\theta \log A < j \leq (1+\varepsilon)\theta \log A} p_j (1 - p_j) \\ Z_{A, \leq} := \frac{1}{\log A} \sum_{j \leq (1-\varepsilon)\theta \log A} p_j (1 - p_j). \end{cases}$$

Combining (3.19), (3.20) and (3.21), we have shown:

$$\begin{cases} Z_{A,>} \xrightarrow{\mathbb{P}} 0 \\ Z_{A,\star} \leq 2\varepsilon\theta + \frac{1}{\log A} \\ Z_{A,\leq} \xrightarrow{\mathbb{P}} \frac{\theta}{6}(1-\varepsilon). \end{cases}$$

Let  $\eta > 0$ . We have

$$\begin{aligned} \mathbb{P}\left(\left|Z_A - \frac{\theta}{6}\right| > \eta\right) &\leq \mathbb{P}\left(Z_{A,>} > \frac{\eta}{4}\right) + \mathbb{P}\left(Z_{A,\star} > \frac{\eta}{4}\right) + \mathbb{P}\left(\left|Z_{A,\leq} - \frac{\theta}{6}(1-\varepsilon)\right| > \frac{\eta}{4}\right) \\ &\quad + \mathbb{P}\left(\frac{\theta}{6}\varepsilon > \frac{\eta}{4}\right) \end{aligned}$$

with

$$\mathbb{P}\left(Z_{A,>} > \frac{\eta}{4}\right) \xrightarrow{A \rightarrow +\infty} 0,$$

$$\mathbb{P}\left(Z_{A,\star} > \frac{\eta}{4}\right) \leq \mathbb{1}_{2\varepsilon\theta + \frac{1}{\log A} > \frac{\eta}{4}},$$

and

$$\mathbb{P}\left(\left|Z_{A,\leq} - \frac{\theta}{6}(1-\varepsilon)\right| > \frac{\eta}{4}\right) \xrightarrow{A \rightarrow +\infty} 0$$

whence taking  $\varepsilon$  sufficiently close to 0 (only depending on  $\eta$  and  $\theta$ , for example  $\varepsilon = \frac{\eta}{12\theta}$  fits well), we get

$$\mathbb{P}\left(\left|Z_A - \frac{\theta}{6}\right| > \eta\right) \xrightarrow{A \rightarrow +\infty} 0,$$

and the proof is complete.

### 3.4 Continuous analog of the Feller coupling

Let  $\mathcal{X}$  be a Poisson process with intensity  $\frac{\theta}{x}dx$  on  $(0, \infty)$ .

In this section we are going to show that one can couple the set of random variables  $\{y_k, k \geq 1\}$  with a set of independent random variables which has the same distribution as  $\mathcal{X} \cap (0, 1)$ , in such a way that these sets are close to each other in  $L^2$ , in a sense which is made precise below.

We choose to label the points of  $\mathcal{X}$  in the following way:

$$0 < \dots < X_3 < X_2 < X_1 < 1 \leq X_0 < X_{-1} < X_{-2} < \dots < \infty. \quad (3.22)$$

For all  $k \in \mathbb{Z}$ , set  $Y_k := X_{k-1} - X_k$ .

Denote  $\mathcal{V} := \{1 - X_1, X_1 - X_2, X_2 - X_3, \dots\}$  and  $\mathcal{W} := \{Y_k : k \in \mathbb{Z}, Y_k < 1\}$ .

To begin with, note that we have the equalities in law  $\{y_k, k \geq 1\} \stackrel{d}{=} \mathcal{V}$  and  $\mathcal{W} \stackrel{d}{=} \mathcal{X} \cap (0, 1)$ . Indeed, this is a direct consequence of the two following lemmas:

**Lemma 3.4.1.**

$$(y_1, y_2, y_3, \dots) \stackrel{d}{=} (1 - X_1, X_1 - X_2, X_2 - X_3, \dots).$$

**Lemma 3.4.2** (Scale invariant spacing lemma).

$$\{Y_k, k \in \mathbb{Z}\} \stackrel{d}{=} \{X_k, k \in \mathbb{Z}\}.$$

We refer to [2] for a proof of Lemma 3.4.1, and [5] for a proof of Lemma 3.4.2. As mentioned by Arratia in [2], the scale-invariant Poisson process  $\mathcal{X}$  is a continuum analog of the sequence  $(\xi_j)_{j \geq 1}$  of independent Bernoulli variables involved in the Feller coupling for generating permutations (see *e.g.* [4] for a description of the Feller coupling and related results). Indeed, for  $j \geq 1$ , the numbers of  $j$ -spacings between two consecutive ones in the infinite word  $\xi_1 \xi_2 \dots$  are independent, and similarly by Lemma 3.4.2 the spacings obtained from the process  $\mathcal{X}$  also form an independent process (in the sense that the numbers of points on disjoint intervals are independent).

Now, we show that the sets  $\mathcal{V}$  and  $\mathcal{W}$  are close from each other in the following sense:

**Lemma 3.4.3.** *There exists a constant number  $C(\theta)$  such that*

$$\mathbb{E}((\#\mathcal{V} \Delta \mathcal{W})^2) \leq C(\theta).$$

*In particular, for all measurable functions  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,*

$$\mathbb{E} \left( \left( \sum_{x \in \mathcal{V}} f(x) - \sum_{x \in \mathcal{W}} f(x) \right)^2 \right) \leq C(\theta) \|f\|_\infty^2.$$

*Proof.* We write

$$\begin{aligned} \mathcal{V} &= \{1 - X_1\} \cup (\{Y_k, k \in \mathbb{Z}\} \setminus \{Y_k, k \leq 1\}) \\ &= \{1 - X_1\} \cup (\mathcal{W} \setminus \{Y_k : k \leq 1, Y_k < 1\}). \end{aligned}$$

Thus,

$$\mathcal{V} \Delta \mathcal{W} \subseteq \{1 - X_1\} \cup \{Y_k : k \leq 1, Y_k < 1\}$$

and then it suffices to show that the number of points in  $\{Y_k : k \leq 0, Y_k < 1\}$  is square-integrable. We write

$$\begin{aligned} &\mathbb{E}((\#\{Y_k : k \leq 0, Y_k < 1\})^2) \\ &= \sum_{k=0, -1, -2, \dots} \mathbb{P}(Y_k < 1) + 2 \sum_{k=-1, -2, \dots} \sum_{\ell=0, -1, \dots, k+1} \mathbb{P}(Y_\ell < 1, Y_k < 1). \end{aligned}$$

For all  $k \leq 0$  and all  $x \in (0, \infty)$ ,

$$\begin{aligned} \mathbb{P}(Y_k \geq x) &= \int_1^{+\infty} \mathbb{P}(Y_k \geq x \mid X_k = s) f_{X_k}(s) ds \\ &= \int_1^{+\infty} \exp \left( - \int_s^{s+x} \frac{\theta}{t} dt \right) f_{X_k}(s) ds \end{aligned}$$

where, using basic properties of Poisson processes,  $f_{X_k}$  (the density function of  $X_k$ ) is



given by

$$\forall s \geq 1, f_{X_k}(s) = \frac{\Lambda(s)^{-k}}{(-k)!} \Lambda'(s) e^{-\Lambda(s)},$$

with  $\Lambda(s) := \int_1^s \frac{\theta}{y} dy$ . Consequently,

$$\begin{aligned} \sum_{k=0,-1,-2,\dots} \mathbb{P}(Y_k < 1) &= \int_1^{+\infty} \left( 1 - \exp \left( - \int_s^{s+1} \frac{\theta}{t} dt \right) \right) \frac{\theta}{s} ds \\ &= \int_1^{+\infty} \left( 1 - \left( \frac{s}{s+1} \right)^\theta \right) \frac{\theta}{s} ds < +\infty \end{aligned} \quad (3.23)$$

since  $1 - \left( \frac{s}{s+1} \right)^\theta \underset{s \rightarrow \infty}{\sim} \frac{\theta}{s}$ .

*Remark 3.4.4.*  $\sum_{k \leq 0} f_{X_k}(s)$  is the density probability function of having a point of the Poisson process at  $s$ , which directly gives  $\theta/s$ .

Now, for all  $k, \ell$  such that  $0 \geq \ell > k$ , denoting by  $f_{(X_\ell, X_k)}$  the density function of the couple  $(X_\ell, X_k)$ ,

$$\begin{aligned} \mathbb{P}(Y_\ell < 1, Y_k < 1) &= \int_{s=1}^{+\infty} \int_{t=s}^{+\infty} \mathbb{P}(Y_\ell \leq 1, Y_k \leq 1 \mid (X_\ell, X_k) = (s, t)) f_{(X_\ell, X_k)}(s, t) dt ds, \end{aligned}$$

where  $\mathbb{P}(Y_\ell \leq 1, Y_k \leq 1 \mid (X_\ell, X_k) = (s, t))$  is equal to the probability that there exist at least one point of the Poisson process in the interval  $(s, s+1]$  and at least one point in the interval  $(t, t+1]$ , that we will denote by  $A_1((s, s+1], (t, t+1])$ . Moreover, the numbers of points of every Poisson process in disjoint intervals are independent. Thus, denoting  $A_j(J)$  the probability that there exists at least  $j$  points of the Poisson process in the interval  $J$ ,

$$\begin{aligned} \mathbb{P}(Y_\ell < 1, Y_k < 1) &= \int_{s=1}^{+\infty} \int_{t=s}^{s+1} A_1((s, s+1], (t, t+1]) f_{(X_\ell, X_k)}(s, t) dt ds \\ &\quad + \int_{s=1}^{+\infty} \int_{t=s+1}^{+\infty} A_1((s, s+1]) A_1((t, t+1]) f_{(X_\ell, X_k)}(s, t) dt ds. \end{aligned}$$

Let us compute an explicit expression for  $f_{(X_\ell, X_k)}$ . For  $x, y > 1$ ,

$$\begin{aligned} \mathbb{P}(X_\ell \leq x, X_k \leq y) &= \int_1^{+\infty} \mathbb{P}(X_\ell \leq x, X_k \leq y \mid X_\ell = s) f_{X_\ell}(s) ds \\ &= \int_1^x \mathbb{P}(X_k \leq y \mid X_\ell = s) f_{X_\ell}(s) ds \\ &= \int_1^x A_{\ell-k}((s, y]) f_{X_\ell}(s) ds. \end{aligned}$$

Thus, for  $x < y$ ,

$$\begin{aligned}
 & \frac{\partial^2}{\partial x \partial y} (\mathbb{P}(X_\ell \leq x, X_k \leq y)) \\
 &= \frac{\partial}{\partial y} (A_{\ell-k}((x, y]) f_{X_\ell}(x)) \\
 &= f_{X_\ell}(x) \frac{\partial}{\partial y} \left( 1 - \sum_{m=0}^{\ell-k-1} \frac{(\int_x^y \frac{\theta}{t} dt)^m}{m!} \exp\left(-\int_x^y \frac{\theta}{t} dt\right) \right) \\
 &= f_{X_\ell}(x) \frac{\theta}{y} \frac{(\int_x^y \frac{\theta}{t} dt)^{\ell-k-1}}{(\ell-k-1)!} \exp\left(-\int_x^y \frac{\theta}{t} dt\right) \\
 &= \frac{\theta^2}{xy} \exp\left(-\int_1^y \frac{\theta}{t} dt\right) \frac{1}{(-k-1)!} \binom{-k-1}{-\ell} \left(\int_1^x \frac{\theta}{t} dt\right)^{-\ell} \left(\int_x^y \frac{\theta}{t} dt\right)^{-k-1-(-\ell)}.
 \end{aligned}$$

Hence

$$\sum_{k=-1, -2, \dots} \sum_{\ell=0, -1, \dots, k+1} f_{(X_\ell, X_k)}(x, y) = \frac{\theta^2}{xy}.$$

*Remark 3.4.5.* This sum is the density probability function of having points of the Poisson process simultaneously at  $x$  and  $y$ , which corresponds to the product of intensities  $\frac{\theta}{x} \times \frac{\theta}{y}$ .

We deduce

$$\begin{aligned}
 & \sum_{k=-1, -2, \dots} \sum_{\ell=0, -1, \dots, k+1} \int_{s=1}^{+\infty} \int_{t=s}^{s+1} A_1((s, s+1], (t, t+1]) f_{(X_\ell, X_k)}(s, t) dt ds \\
 &= \int_{s=1}^{+\infty} \int_{t=s}^{s+1} \frac{\theta^2}{st} dt ds \\
 &\leq \theta^2 \int_{s=1}^{+\infty} \frac{1}{s^2} ds = \theta^2 < +\infty
 \end{aligned}$$

and

$$\begin{aligned}
 & \sum_{k=-1, -2, \dots} \sum_{\ell=0, -1, \dots, k+1} \int_{s=1}^{+\infty} \int_{t=s+1}^{+\infty} A_1((s, s+1]) A_1((t, t+1]) f_{(X_\ell, X_k)}(s, t) dt ds \\
 &= \int_{s=1}^{+\infty} \int_{t=s+1}^{+\infty} A_1((s, s+1]) A_1((t, t+1]) \frac{\theta^2}{st} dt ds \\
 &\leq \left( \int_{s=1}^{+\infty} A_1((s, s+1]) \frac{\theta}{s} ds \right)^2 < +\infty
 \end{aligned}$$

by (3.23). Consequently,

$$\sum_{k=-1, -2, \dots} \sum_{\ell=0, -1, \dots, k+1} \mathbb{P}(Y_\ell < 1, Y_k < 1) < +\infty.$$

This shows the first part of the lemma. The second part of the lemma immediately

derives from the first part and the classical inequalities

$$\left| \sum_{x \in \mathcal{V}} f(x) - \sum_{x \in \mathcal{W}} f(x) \right| \leq \sum_{x \in \mathcal{V} \Delta \mathcal{W}} |f(x)| \leq \|f\|_\infty \# \mathcal{V} \Delta \mathcal{W}.$$

□

A key result for proving Theorem 3.1.7 is the following simple version of the Campbell's theorem:

**Theorem 3.4.6** (Campbell). *Let  $N$  be a Poisson process with intensity  $\Lambda$  on  $\mathbb{R}$ . Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a measurable function, and denote  $T := \sum_{x \in N} f(x)$ . Assume  $\int_{\mathbb{R}} \min(|f(x)|, 1) \Lambda(dx) < +\infty$ . Then for all real numbers  $t$ ,*

$$\mathbb{E}(e^{itT}) = \exp \left( \int_{\mathbb{R}} (e^{itf(x)} - 1) \Lambda(dx) \right).$$

Moreover,

$$\mathbb{E}(T) = \int_{\mathbb{R}} f(x) \Lambda(dx)$$

and

$$\text{Var}(T) = \int_{\mathbb{R}} f(x)^2 \Lambda(dx)$$

if these integrals converge.

We refer to [61] for a proof of Theorem 3.4.6.

From Campbell's theorem, we deduce the following lemma, which will be useful in the next section.

**Lemma 3.4.7.** *For all  $u \in \mathbb{R}_+^*$ , let  $f_u$  be function from  $\mathbb{R}$  to  $\mathbb{R}$ , and let  $T_u := \sum_{y \in \mathcal{X} \cap (0,1)} f_u(y)$ . We assume that the following conditions are satisfied:*

- *There exists  $K > 0$  such that for all  $u$ ,  $|f_u| \leq K$ .*
- *For all  $u$ ,  $\int_0^1 |f_u(x)| \frac{\theta}{x} dx < +\infty$ .*
- *$\int_0^1 f_u(x)^2 \frac{\theta}{x} dx \xrightarrow{u \rightarrow +\infty} +\infty$ .*

Then as  $u \rightarrow +\infty$ ,

$$\frac{T_u - \mathbb{E}(T_u)}{\sqrt{\text{Var}(T_u)}} \xrightarrow{d} \mathcal{N}(0, 1).$$

*Proof.* Denote  $N_u := \frac{T_u - \mathbb{E}(T_u)}{\sqrt{\text{Var}(T_u)}}$ , and  $v_u := \sqrt{\text{Var}(T_u)}$ . By Theorem 3.4.6 the Fourier transform of  $N_u$  is given by

$$\begin{aligned} \forall t \in \mathbb{R}, \quad \mathbb{E}(e^{itN_u}) &= \exp \left( \int_0^1 \left( e^{i \frac{t}{v_u} f_u(x)} - 1 \right) \frac{\theta}{x} dx \right) e^{-i \frac{t}{v_u} \mathbb{E}(T_u)} \\ &= \exp \left( -\frac{t^2}{2} + \int_0^1 \left( \sum_{k=3}^{+\infty} \frac{1}{k!} \left( i \frac{t}{v_u} f_u(x) \right)^k \right) \frac{\theta}{x} dx \right) \end{aligned}$$

with

$$\begin{aligned} \left| \sum_{k=3}^{+\infty} \frac{1}{k!} \left( i \frac{t}{v_u} f_u(x) \right)^k \right| &\leq \sum_{k=3}^{+\infty} \frac{|t|^k K^{k-2} f_u(x)^2}{k! v_u^k} \\ &\leq \frac{|t|^3 K f_u(x)^2}{v_u^3} \exp \left( \frac{|t|K}{v_u} \right). \end{aligned}$$

Thus

$$\begin{aligned} \int_0^1 \left| \sum_{k=3}^{+\infty} \frac{1}{k!} \left( i \frac{t}{v_u} f_u(x) \right)^k \right| \frac{\theta}{x} dx &\leq \exp \left( \frac{|t|K}{v_u} \right) \frac{|t|^3 K}{v_u^3} \int_0^1 f_u(x)^2 \frac{\theta}{x} dx \\ &= \exp \left( \frac{|t|K}{v_u} \right) \frac{|t|^3 K}{v_u} \\ &\underset{u \rightarrow +\infty}{=} \exp(o(1)) o(1) = o(1) \end{aligned}$$

and finally

$$\begin{aligned} \mathbb{E} (e^{itN_u}) &\underset{u \rightarrow +\infty}{=} \exp \left( -\frac{t^2}{2} \right) \exp(o(1)) \\ &= \exp \left( -\frac{t^2}{2} \right) + o(1). \end{aligned}$$

□

### 3.5 Limiting point process related to permutation matrices. Proof of Theorem 3.1.7

Let us introduce the random variable  $X(s, t)$  which counts the number of points, between the positive real numbers  $s$  and  $t$ , of the limiting point process related to permutation matrices (without modification), i.e

$$X(s, t) = t - s - \sum_{j=1}^{+\infty} (\{ty_j\} - \{sy_j\}).$$

Here, we choose to generate the ensemble  $\{y_j, j \geq 1\}$  using the continuous analog of the Feller coupling described above.

Let us begin with three lemmas before stating results about  $X(s, t)$ . Since their proofs are technical we postpone them in Appendix.

**Lemma 3.5.1.** *Let  $n$  be a positive integer.*

•

$$\int_0^1 \frac{\{nx\}}{x} dx \underset{n \rightarrow \infty}{=} \frac{1}{2} \log n + \mathcal{O}(1).$$

•

$$\int_0^1 \{nx\} \log x dx \underset{n \rightarrow \infty}{=} -\frac{1}{2} + \frac{1}{12n} \log n + \mathcal{O} \left( \frac{1}{n} \right).$$

**Lemma 3.5.2.** *Let  $\ell \in \mathbb{N}^*$ . Then*

$$\begin{aligned} \sum_{k=1}^{n-1} \left( 2 \left\{ \ell \frac{k}{n} \right\} - 1 \right) \log \frac{k}{n} &\underset{n \rightarrow \infty}{=} \left[ \frac{\ell}{2} + 2 \sum_{m=1}^{\ell-1} \frac{m}{\ell} \log \frac{m}{\ell} \right] n - \frac{1}{2} \log n + \mathcal{O}(1) \\ \sum_{k=1}^{n-1} \left( 2 \left\{ -\ell \frac{k}{n} \right\} - 1 \right) \log \frac{k}{n} &\underset{n \rightarrow \infty}{=} - \left[ \frac{\ell}{2} + 2 \sum_{m=1}^{\ell-1} \frac{m}{\ell} \log \frac{m}{\ell} \right] n + \frac{1}{2} \log n + \mathcal{O}(1) \end{aligned}$$

**Lemma 3.5.3.** *Let  $p, q$  be two positive integers. Then*

$$\begin{aligned} &\int_0^1 \frac{(\{px\} - \{qx\})^2}{x} dx \\ &= -2(p-q) \int_0^1 (\{px\} - \{qx\}) \log x dx - \sum_{k=1}^{p-1} \left( 2 \left\{ q \frac{k}{p} \right\} - 1 \right) \log \left( \frac{k}{p} \right) \\ &\quad - \sum_{j=1}^{q-1} \left( 2 \left\{ p \frac{j}{q} \right\} - 1 \right) \log \left( \frac{j}{q} \right) - 2 \sum_{m=1}^{\gcd(p,q)-1} \log \left( \frac{m}{\gcd(p,q)} \right). \end{aligned}$$

Let  $a, b > 0$ . Define  $f_{a,b} : x \mapsto \{(a+b)x\} - \{ax\}$ , and denote  $S := \sum_{y \in \mathcal{V}} f_{a,b}(y) = b - X(a, a+b)$  and  $T := \sum_{y \in \mathcal{W}} f_{a,b}(y)$ .

### 3.5.1 Approximation of $S$ by $T$

Using Lemma 3.4.3,

$$|\mathbb{E}(S) - \mathbb{E}(T)| \leq \mathbb{E}|S - T| \leq \|f_{a,b}\|_{\infty} C(\theta) \leq C(\theta) \quad (3.24)$$

and

$$|\sqrt{\text{Var}(S)} - \sqrt{\text{Var}(T)}| \leq \sqrt{\text{Var}(S - T)} \leq \sqrt{\mathbb{E}((S - T)^2)} \leq \|f_{a,b}\|_{\infty} \sqrt{C(\theta)} \leq \sqrt{C(\theta)}. \quad (3.25)$$

Therefore, as soon as  $\text{Var}(T) \rightarrow +\infty$  we will get

$$\frac{\mathbb{E}(S) - S}{\sqrt{\text{Var}(S)}} - \frac{\mathbb{E}(T) - T}{\sqrt{\text{Var}(T)}} \xrightarrow{\mathbb{P}} 0.$$

Moreover, it is easy to check that  $\int_0^1 \frac{|f_{a,b}(x)|}{x} dx < +\infty$ , as for all  $x \in \left(0, \frac{1}{a+b}\right)$  we have  $f_{a,b}(x) = bx$ . We deduce by Lemma 3.4.7 and Slutsky's theorem that as soon as  $\text{Var}(T) \rightarrow +\infty$ , we have  $\frac{\mathbb{E}(S) - S}{\sqrt{\text{Var}(S)}} \xrightarrow{d} \mathcal{N}(0, 1)$ , i.e.  $\frac{X(a, a+b) - \mathbb{E}(X(a, a+b))}{\sqrt{\text{Var}(X(a, a+b))}} \xrightarrow{d} \mathcal{N}(0, 1)$ . Furthermore, Theorem 3.4.6 applies and gives

$$\mathbb{E}(T) = \theta \int_0^1 \frac{f_{a,b}(x)}{x} dx$$

and

$$\text{Var}(T) = \theta \int_0^1 \frac{f_{a,b}(x)^2}{x} dx.$$

### 3.5.2 Proof of Theorem 3.1.7 (i)

**Proof for  $a, b \in \mathbb{N}^*$**

Assume  $a, b \in \mathbb{N}^*$ . Then Lemmas 3.5.1, 3.5.2 and 3.5.3 provides all we need for the computation of the asymptotics of  $\mathbb{E}(T)$  and  $\text{Var}(T)$  when  $b$  tends to infinity.

Denote  $p = a$  and  $q = a + b$ . If  $a$  is fixed and  $b$  goes to infinity, then using Lemma 3.5.1,

$$\mathbb{E}(T) = \frac{\theta}{2} \log b + \mathcal{O}_\theta(1) \quad (3.26)$$

and

$$\begin{aligned} & -2(p-q) \int_0^1 (\{px\} - \{qx\}) \log x dx \\ &= 2(q-p) \left( \int_0^1 \{px\} \log x dx + \frac{1}{2} - \frac{1}{12} \frac{\log q}{q} + \mathcal{O}\left(\frac{1}{q}\right) \right) \\ &= \frac{q-p}{q} \left( \left(1 + 2 \int_0^1 \{px\} \log x dx\right) q - \frac{1}{6} \log q + \mathcal{O}(1) \right) \end{aligned}$$

with  $\frac{q-p}{q} = 1 + \mathcal{O}(\frac{1}{q})$ . Besides, clearly since  $p$  is fixed

$$\sum_{k=1}^{p-1} \left( 2 \left\{ k \frac{q}{p} \right\} - 1 \right) \log \left( \frac{k}{p} \right) = \mathcal{O}(1),$$

and using Lemma 3.5.2,

$$\sum_{j=1}^{q-1} \left( 2 \left\{ p \frac{j}{q} \right\} - 1 \right) \log \left( \frac{j}{q} \right) = \left[ \frac{p}{2} + 2 \sum_{m=1}^{p-1} \frac{m}{p} \log \frac{m}{p} \right] q - \frac{1}{2} \log q + \mathcal{O}(1).$$

Furthermore, as  $\gcd(p, q) \leq p$ ,

$$\sum_{m=1}^{\gcd(p,q)-1} \log \left( \frac{m}{\gcd(p,q)} \right) = \mathcal{O}(1).$$

We deduce by Lemma 3.5.3

$$\begin{aligned} \int_0^1 \frac{(\{px\} - \{qx\})^2}{x} dx &= \left( 1 + 2 \int_0^1 \{px\} \log x dx - \frac{p}{2} - 2 \sum_{m=1}^{p-1} \frac{m}{p} \log \frac{m}{p} \right) q \\ &\quad + \left( -\frac{1}{6} + \frac{1}{2} \right) \log q + \mathcal{O}(1) \\ &= \frac{1}{3} \log q + \mathcal{O}(1). \end{aligned}$$

Indeed,

$$\begin{aligned}
 \int_0^1 \{px\} \log x dx &= \sum_{k=0}^{p-1} \int_{\frac{k}{p}}^{\frac{k+1}{p}} (px - k) \log x dx \\
 &= p \int_0^1 x \log x dx \\
 &\quad - \sum_{k=1}^{p-1} \left( \sum_{j=1}^k 1 \right) \left( \left( \frac{k+1}{p} \log \frac{k+1}{p} - \frac{k+1}{p} \right) - \left( \frac{k}{p} \log \frac{k}{p} - \frac{k}{p} \right) \right) \\
 &= -\frac{p}{4} + \sum_{j=1}^p \left( \frac{j}{p} \log \frac{j}{p} - \frac{j}{p} \right) + p \\
 &= \frac{p}{4} + \sum_{j=1}^{p-1} \frac{j}{p} \log \frac{j}{p} - \frac{1}{2}.
 \end{aligned}$$

It follows

$$\text{Var}(T) = \frac{\theta}{3} \log b + \mathcal{O}_\theta(1), \quad (3.27)$$

and as  $\left| \sqrt{\text{Var}(T)} - \sqrt{\text{Var}(S)} \right| = \mathcal{O}_\theta(1)$ , then

$$\begin{aligned}
 \text{Var}(X(a, a+b)) &= \text{Var}(S) = \left( \left( \frac{\theta}{3} \log b + \mathcal{O}_\theta(1) \right)^{1/2} + \mathcal{O}_\theta(1) \right)^2 \\
 &= \frac{\theta}{3} \log b + \mathcal{O}_\theta(\sqrt{\log b}).
 \end{aligned}$$

From the previous paragraph we deduce

$$\frac{X(a, a+b) - \mathbb{E}(X(a, a+b))}{\sqrt{\text{Var}(X(a, a+b))}} \xrightarrow[b \rightarrow \infty]{d} \mathcal{N}(0, 1). \quad (3.28)$$

### Generalization for all $a, b$

Assume now  $a, b$  to be positive real numbers, with  $b > 1$ . We have the inequalities

$$\begin{aligned}
 0 \leq X(a, a+b) - X(\lceil a \rceil, \lfloor a \rfloor + \lfloor b \rfloor) &= X(a, \lceil a \rceil) + X(\lfloor a \rfloor + \lfloor b \rfloor, a+b) \\
 &\leq X(a, \lceil a \rceil) + X(\lfloor a \rfloor + \lfloor b \rfloor, \lfloor a \rfloor + 2 + \lfloor b \rfloor),
 \end{aligned}$$

with

$$\mathbb{E}(X(\lfloor a \rfloor + \lfloor b \rfloor, \lfloor a \rfloor + 2 + \lfloor b \rfloor)) \xrightarrow[b \rightarrow \infty]{} \mathcal{O}_\theta(1) \quad (3.29)$$

by (3.24) and Lemma 3.5.1. Moreover,

$$\begin{aligned}
 & \left| \sqrt{\text{Var}(X(a, a+b))} - \sqrt{\text{Var}(X(\lceil a \rceil, \lfloor a \rfloor + \lfloor b \rfloor))} \right| \\
 & \leq \sqrt{\text{Var}(X(a, a+b) - X(\lceil a \rceil, \lfloor a \rfloor + \lfloor b \rfloor))} \\
 & = \sqrt{\text{Var}(X(a, \lceil a \rceil) + X(\lfloor a \rfloor + \lfloor b \rfloor, a+b))} \\
 & \leq \sqrt{3} \sqrt{\text{Var}(X(a, \lceil a \rceil)) + \text{Var}(X(\lfloor a \rfloor + \lfloor b \rfloor, \lfloor a \rfloor + \lfloor b \rfloor)) + \text{Var}(X(\lfloor a \rfloor + \lfloor b \rfloor, a+b))} \\
 & \leq 3 + \text{Var}(X(a, \lceil a \rceil)) + \text{Var}(X(\lfloor a \rfloor + \lfloor b \rfloor, \lfloor a \rfloor + \lfloor b \rfloor)) + \text{Var}(X(\lfloor a \rfloor + \lfloor b \rfloor, a+b)).
 \end{aligned}$$

Let us show that  $\text{Var}(X(\lfloor a \rfloor + \lfloor b \rfloor, \lfloor a \rfloor + \lfloor b \rfloor)) + \text{Var}(X(\lfloor a \rfloor + \lfloor b \rfloor, a+b)) = \mathcal{O}_\theta(1)$ .

For the first term, from (3.25) it is enough to show

$$\int_0^1 \frac{(\{(\lfloor a \rfloor + \lfloor b \rfloor)x\} - \{(\lfloor a \rfloor + \lfloor b \rfloor)x\})^2}{x} dx \underset{b \rightarrow \infty}{=} \mathcal{O}(1). \quad (3.30)$$

For the sake of simplicity, denote  $m = \lfloor a \rfloor$ . We have, for all  $x \in [0, 1]$ ,

$$\begin{aligned}
 \{(m+b)x\} - \{(m+\lfloor b \rfloor)x\} &= \{(m+b)x - (m+\lfloor b \rfloor)x\} \\
 &= \mathbb{1}_{\{(m+b)x - (m+\lfloor b \rfloor)x + \{(m+\lfloor b \rfloor)x\} > 1\}} \\
 &= \{b\}x - \mathbb{1}_{\{b\}x + \{(m+\lfloor b \rfloor)x\} > 1}
 \end{aligned}$$

so that

$$\int_0^1 \frac{(\{(\lfloor a \rfloor + \lfloor b \rfloor)x\} - \{(\lfloor a \rfloor + \lfloor b \rfloor)x\})^2}{x} dx \underset{b \rightarrow \infty}{=} I_b + \mathcal{O}(1)$$

where  $I_b := \int_0^1 \frac{1}{x} \mathbb{1}_{\{b\}x + \{(m+\lfloor b \rfloor)x\} > 1} dx$ . We want to show  $I_b = \mathcal{O}(1)$ . We cut the integral as follows:

$$\begin{aligned}
 I_b &= \int_0^{1/(m+\lfloor b \rfloor)} \frac{1}{x} \mathbb{1}_{\{b\}x + \{(m+\lfloor b \rfloor)x\} > 1} dx \\
 &+ \sum_{k=1}^{m+\lfloor b \rfloor-1} \int_{k/(m+\lfloor b \rfloor)}^{(k+1)/(m+\lfloor b \rfloor)} \frac{1}{x} \mathbb{1}_{\{b\}x + \{(m+\lfloor b \rfloor)x\} > 1} dx.
 \end{aligned}$$

We have

$$\begin{aligned}
 \int_0^{1/(m+\lfloor b \rfloor)} \frac{1}{x} \mathbb{1}_{\{b\}x + \{(m+\lfloor b \rfloor)x\} > 1} dx &= \int_0^{1/(m+\lfloor b \rfloor)} \frac{1}{x} \mathbb{1}_{(m+b)x > 1} dx \\
 &= \int_{1/(m+b)}^{1/(m+\lfloor b \rfloor)} \frac{1}{x} dx = \log \left( \frac{m+b}{m+\lfloor b \rfloor} \right) \\
 &\xrightarrow{b \rightarrow \infty} 0,
 \end{aligned}$$



and for all  $k \geq 1$ ,

$$\begin{aligned}
 \int_{k/(m+\lfloor b \rfloor)}^{(k+1)/(m+\lfloor b \rfloor)} \frac{1}{x} \mathbb{1}_{\{b\}x + \{(m+\lfloor b \rfloor)x\} > 1} dx &= \int_{k/(m+\lfloor b \rfloor)}^{(k+1)/(m+\lfloor b \rfloor)} \frac{1}{x} \mathbb{1}_{x > (k+1)/(m+b)} dx \\
 &\leq \frac{1}{k/(m+\lfloor b \rfloor)} \int_{(k+1)/(m+b)}^{(k+1)/(m+\lfloor b \rfloor)} dx \\
 &= \frac{k+1}{k} (m+\lfloor b \rfloor) \left( \frac{1}{m+\lfloor b \rfloor} - \frac{1}{m+b} \right) \\
 &\leq \frac{2}{m+\lfloor b \rfloor}.
 \end{aligned}$$

Hence  $I_b \leq o(1) + 2 = \mathcal{O}(1)$ . A very similar computation gives  $\text{Var}(X(\lfloor a \rfloor + b, a+b)) = \mathcal{O}_\theta(1)$ .

We deduce

$$\sqrt{\text{Var}(X(a, a+b))} = \sqrt{\text{Var}(X(\lceil a \rceil, \lfloor a \rfloor + \lfloor b \rfloor))} + \mathcal{O}_\theta(1)$$

which, combining with (3.25), yields

$$\sqrt{\text{Var}(X(a, a+b))} = \sqrt{\frac{\theta}{3} \log b} + \mathcal{O}_\theta(1). \quad (3.31)$$

Using Markov inequality, (3.29) and (3.31) imply

$$\frac{X(a, a+b) - X(\lceil a \rceil, \lfloor a \rfloor + \lfloor b \rfloor)}{\sqrt{\text{Var}(X(a, a+b))}} \xrightarrow[b \rightarrow \infty]{\mathbb{P}} 0$$

and applying Slutsky's lemma it follows

$$\frac{X(a, a+b) - \mathbb{E}(X(a, a+b))}{\sqrt{\text{Var}(X(a, a+b))}} \xrightarrow[b \rightarrow \infty]{d} \mathcal{N}(0, 1)$$

which completes the proof.

### 3.5.3 Proof of Theorem 3.1.7 (ii)

#### Case $\nu$ rational

**Lemma 3.5.4.** *Let  $f$  be a non-negative function on  $[0, 1]$  such that  $f$  is integrable on  $[0, 1]$  and  $x \mapsto \frac{f(x)}{x}$  is integrable in the neighbourhood of 0. Let  $t \in \mathbb{R}$ . Then*

$$\int_0^1 \frac{f(\{tx\})}{x} dx \underset{t \rightarrow +\infty}{=} (\log t) \int_0^1 f(x) dx + \mathcal{O}(1).$$

*Proof.* Let  $t \geq 2$ . It suffices to write

$$\int_0^1 \frac{f(\{tx\})}{x} dx = \int_0^t \frac{f(\{x\})}{x} dx = \int_0^1 \frac{f(x)}{x} dx + \sum_{k=1}^{\lfloor t \rfloor - 1} \int_k^{k+1} \frac{f(\{x\})}{x} dx + \int_{\lfloor t \rfloor}^t \frac{f(\{x\})}{x} dx$$

and to notice that

$$\sum_{k=1}^{\lfloor t \rfloor - 1} \frac{1}{k+1} \int_0^1 f(x) dx \leq \sum_{k=1}^{\lfloor t \rfloor - 1} \int_k^{k+1} \frac{f(\{x\})}{x} dx \leq \sum_{k=1}^{\lfloor t \rfloor - 1} \frac{1}{k} \int_0^1 f(x) dx$$

and

$$\int_{\lfloor t \rfloor}^t \frac{f(\{x\})}{x} dx \leq \frac{1}{\lfloor t \rfloor} \int_0^1 f(x) dx.$$

□

We are ready to prove point (ii) of the theorem for the case  $\nu = \frac{r}{s}$  with  $\frac{r}{s} > 1$  and  $\gcd(r, s) = 1$ . Let  $a \in \mathbb{R}$ . We want to show

$$\frac{X(a, \frac{r}{s}a) - (\frac{r}{s} - 1)a}{\sqrt{\theta \log a (\frac{1}{6} - \frac{1}{6sr})}} \xrightarrow{a \rightarrow +\infty} \mathcal{N}(0, 1).$$

With the notation  $T = \sum_{y \in \mathcal{W}} \{(a+b)y\} - \{ay\}$ , we established that as soon as  $\text{Var}(T) \rightarrow \infty$ ,

$$\frac{X(a, a+b) - b + \mathbb{E}(T)}{\sqrt{\text{Var}(T)}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Set  $b = (\frac{r}{s} - 1)a$ . Using twice Lemma 3.5.4 with the identity function and  $t = \frac{r}{s}a$  and then  $t = a$ , we get by subtraction

$$\mathbb{E}(T) = \theta \left( \frac{1}{2} \log \left( \frac{r}{s}a \right) - \frac{1}{2} \log a + \mathcal{O}(1) \right) = \mathcal{O}_\theta(1).$$

Now, denoting  $t := \frac{a}{s}$ , for all  $x \in [0, 1]$ ,

$$\{ax\} - \left\{ \frac{r}{s}ax \right\} = \{stx\} - \{rtx\} = (s-r)\{tx\} - \sum_{m=1}^{s-1} \mathbb{1}_{\{tx\} \geq \frac{m}{s}} + \sum_{n=1}^{r-1} \mathbb{1}_{\{tx\} \geq \frac{n}{r}}.$$

Hence applying Lemma 3.5.4 with the function

$$f : x \mapsto \left( (s-r)x - \sum_{m=1}^{s-1} \mathbb{1}_{x \geq \frac{m}{s}} + \sum_{n=1}^{r-1} \mathbb{1}_{x \geq \frac{n}{r}} \right)^2,$$

we get

$$\int_0^1 \frac{(\{ax\} - \{\frac{r}{s}ax\})^2}{x} dx = \int_0^1 \frac{f(\{tx\})}{x} dx \stackrel{t \rightarrow +\infty}{=} \log t \int_0^1 f(x) dx + \mathcal{O}(1).$$

In Chapter 2 Appendix B, we shew that

$$\int_0^1 f(x) dx = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n (\{sj\alpha\} - \{rj\alpha\})^2$$

where  $\alpha$  is any arbitrary irrational number, and we computed this limit explicitly, equal to  $\frac{1}{6} - \frac{1}{6sr}$ , which gives the claim.

### Case $\nu$ irrational

Let  $\nu$  be an irrational number. For all  $a > 0$ , let  $\mu_a$  be the empirical measure of  $(U_a, \nu U_a)$  on  $(\mathbb{R}/\mathbb{Z})^2$ , where  $U_a$  is a uniform random variable on  $[0, a]$ .

Then, the Fourier transform of  $\mu_a$  is given for all  $(k, l) \in \mathbb{Z}^2$  by

$$\widehat{\mu_a}(k, l) = \frac{1}{a} \int_0^a e^{2i\pi(k+l\nu)x} dx \xrightarrow{a \rightarrow \infty} \mathbb{1}_{k+l\nu=0}.$$

Since  $\nu$  is irrational, then  $k + l\nu = 0$  if and only if  $(k, l) = (0, 0)$ . We deduce that  $\mu_a$  converges to the Lebesgue measure of dimension 2 on  $[0, 1]^2$ .

Let  $f$  be a function from  $(\mathbb{R}/\mathbb{Z})^2$  to  $\mathbb{R}$  defined by  $f(x, y) = (x - y)^2$ .  $f$  is continuous everywhere, excepted on  $\mathbb{R}/\mathbb{Z} \times \{\bar{0}\}$  and  $\{\bar{0}\} \times \mathbb{R}/\mathbb{Z}$ , which are of measure zero with respect to the Lebesgue measure of dimension 2. Hence by the continuous mapping theorem,

$$\int f d\mu_a \xrightarrow{a \rightarrow \infty} \int_0^1 \int_0^1 (x - y)^2 dx dy = \frac{1}{6},$$

so that by a change of variables we get

$$\int_0^1 (\{ax\} - \{\nu ax\})^2 dx = \frac{1}{a} \int_0^a (\{x\} - \{\nu x\})^2 dx = \int f d\mu_a \xrightarrow{a \rightarrow \infty} \frac{1}{6}. \quad (3.32)$$

It remains to show that this implies

$$\frac{1}{\log a} \int_0^1 \frac{(\{ax\} - \{\nu ax\})^2}{x} dx \xrightarrow{a \rightarrow \infty} \frac{1}{6}. \quad (3.33)$$

Assume  $a > 1$ . We write

$$\begin{aligned} & \int_0^a \frac{f(\{x\}, \{\nu x\})}{x} dx \\ &= \int_0^1 \frac{f(\{x\}, \{\nu x\})}{x} dx + \sum_{k=1}^{\lfloor a \rfloor - 1} \int_k^{k+1} \frac{f(\{x\}, \{\nu x\})}{x} dx + \int_{\lfloor a \rfloor}^a \frac{f(\{x\}, \{\nu x\})}{x} dx \end{aligned}$$

with  $\int_0^1 \frac{f(\{x\}, \{\nu x\})}{x} dx < +\infty$ , and  $\int_{\lfloor a \rfloor}^a \frac{f(\{x\}, \{\nu x\})}{x} dx = \mathcal{O}(1)$  since  $f$  is bounded. Moreover, for all integers  $k \geq 1$ ,

$$\frac{1}{k+1} \int_k^{k+1} f(\{x\}, \{\nu x\}) dx \leq \int_k^{k+1} \frac{f(\{x\}, \{\nu x\})}{x} dx \leq \frac{1}{k} \int_k^{k+1} f(\{x\}, \{\nu x\}) dx.$$

For the right-hand side inequality, denoting  $a_k := \frac{1}{k}$ , and  $b_k := \int_k^{k+1} f(\{x\}, \{\nu x\}) dx$ , a

summation by parts gives, for all  $n \geq 1$ ,

$$\begin{aligned} \sum_{k=1}^n a_k b_k &= a_n \sum_{k=1}^n b_k - \sum_{k=1}^{n-1} \sum_{m=1}^k b_m (a_{k+1} - a_k) \\ &= \frac{1}{n} \int_1^{n+1} f(\{x\}, \{\nu x\}) dx + \sum_{k=1}^{n-1} \frac{1}{k(k+1)} \int_1^{k+1} f(\{x\}, \{\nu x\}) dx \end{aligned}$$

and from (3.32) we deduce

$$\begin{aligned} \sum_{k=1}^n a_k b_k &= \frac{1}{6} + o(1) + \sum_{k=1}^{n-1} \frac{1}{k} \left( \frac{1}{6} + o(1) \right) \\ &= \frac{1}{6} \log n + o(\log n) \end{aligned}$$

as  $n \rightarrow \infty$ . Replacing  $a_k$  by  $\frac{1}{k+1}$  leads to the same asymptotic expression. Hence, from the squeeze theorem,

$$\frac{1}{\log a} \sum_{k=1}^{\lfloor a \rfloor - 1} \int_k^{k+1} \frac{f(\{x\}, \{\nu x\})}{x} dx \xrightarrow{a \rightarrow \infty} \frac{1}{6},$$

which gives (3.33).

### 3.6 Translation of the limiting point process related to permutation matrices. Proof of Proposition 3.1.9

In this section we show that the translation of the limiting point process related to permutation matrices converges to the limiting point process related to modified permutation matrices. The precise statement corresponds to Proposition 3.1.9. We will need the following lemma:

**Lemma 3.6.1.** *For all  $j \in \mathbb{N}^*$ ,*

$$(\{Ay_1\}, \{Ay_2\}, \dots, \{Ay_j\}, y_1, \dots, y_j) \xrightarrow[A \rightarrow \infty]{d} (\Phi_1, \dots, \Phi_j, y_1, \dots, y_j)$$

where  $\Phi_1, \dots, \Phi_j$  are i.i.d random variables uniformly distributed on  $[0, 1)$  and independent of  $y_1, \dots, y_j$ .

*Proof.* Let  $j \in \mathbb{N}^*$ . We know that  $\vec{y} := (y_1, \dots, y_j)$  has a density with respect to the Lebesgue measure (see [4]). Hence, for all  $\vec{k} \in \mathbb{Z}^j$  and  $\vec{\lambda} \in \mathbb{R}^j$ ,

$$\mathbb{E} \left[ e^{2i\pi A\vec{k} \cdot \vec{y} + i\vec{\lambda} \cdot \vec{y}} \right] = \mathbb{E} \left[ e^{i\vec{y} \cdot (2\pi A\vec{k} + \vec{\lambda})} \right] = \widehat{\mu_{\vec{y}}}(2\pi A\vec{k} + \vec{\lambda}) \xrightarrow[A \rightarrow \infty]{} 0$$

as soon as  $\vec{k} \neq 0$ , applying Riemann-Lebesgue lemma.  $\square$

We are ready to prove Proposition 3.1.9.

Let  $f$  be a continuous function from  $\mathbb{R}$  to  $\mathbb{C}$  such that  $\text{supp} f \subset [-M, M]$  for any  $M > 0$ . With the same notations as in the previous lemma, we want to show:

$$\sum_{\substack{k \in \mathbb{Z} \setminus \{0\} \\ j \geq 1}} f\left(\frac{k}{y_j} - A\right) \xrightarrow[A \rightarrow \infty]{d} \sum_{\substack{k \in \mathbb{Z} \\ j \geq 1}} f\left(\frac{k - \Phi_j}{y_j}\right). \quad (3.34)$$

Let  $j_0 \in \mathbb{N}^*$ . The probability that there exists non-zero terms in the sum  $\sum_{\substack{k \in \mathbb{Z} \\ j \geq j_0}} f\left(\frac{k - \Phi_j}{y_j}\right)$  is

$$\begin{aligned} & \mathbb{P}\left(\exists k \in \mathbb{Z}, \exists j \geq j_0, \left|\frac{k - \Phi_j}{y_j}\right| \leq M\right) \\ & \leq \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} \mathbb{P}\left(\left|\frac{k - \Phi_j}{y_j}\right| \leq M\right) \\ & = \sum_{j \geq j_0} \left[ \mathbb{P}(\Phi_j \leq M y_j) + \mathbb{P}(1 - \Phi_j \leq M y_j) + \sum_{k = -\lfloor M \rfloor - 1}^{-1} \mathbb{P}\left(y_j \geq \frac{|k - \Phi_j|}{M}\right) \right. \\ & \quad \left. + \sum_{k=2}^{\lfloor M \rfloor + 1} \mathbb{P}\left(y_j \geq \frac{|k - \Phi_j|}{M}\right) \right] \\ & \leq \sum_{j \geq j_0} \left[ M r^j + M r^j + \sum_{k = -\lfloor M \rfloor - 1}^{-1} \frac{M r^j}{-k} + \sum_{k=2}^{\lfloor M \rfloor + 1} \frac{M r^j}{k-1} \right] \\ & \leq \frac{2M(1 + \log(M+1))}{1-r} r^{j_0} \end{aligned}$$

where we recall that  $r$  is the constant given by (3.5). Thus

$$\mathbb{P}\left(\exists j \geq j_0, \exists k \in \mathbb{Z}, f\left(\frac{k - \Phi_j}{y_j}\right) \neq 0\right) \xrightarrow{j_0 \rightarrow \infty} 0. \quad (3.35)$$

Let  $A > 2M$ . The probability that there exists non-zero terms in the sum  $\sum_{\substack{k \in \mathbb{Z} \setminus \{0\} \\ j \geq j_0}} f\left(\frac{k}{y_j} - A\right)$  is

$$\begin{aligned} & \mathbb{P}\left(\exists k \in \mathbb{Z} \setminus \{0\}, \exists j \geq j_0, \left|\frac{k}{y_j} - A\right| \leq M\right) \\ & = \mathbb{P}\left(\exists k \in \mathbb{Z} \cap [1, M+A], \exists j \geq j_0, \frac{k}{M+A} \leq y_j \leq \frac{k}{A-M}\right) \end{aligned}$$

since  $0 < y_j < 1$  a.s. for all  $j$ . Moreover, the intervals  $\left[\frac{k}{M+A}, \frac{k}{A-M}\right]$  and  $\left[\frac{k+1}{M+A}, \frac{k+1}{A-M}\right]$  overlap if and only if  $k \geq \frac{A-M}{2M}$ .

On the one hand,

$$\begin{aligned}
 & \mathbb{P} \left( \exists k \in \mathbb{Z} \cap \left[ \frac{A-M}{2M}, M+A \right], \exists j \geq j_0, \frac{k}{M+A} \leq y_j \leq \frac{k}{A-M} \right) \\
 & \leq \mathbb{P} \left( \exists j \geq j_0, y_j \geq \frac{\frac{A-M}{2M}}{M+A} \right) \\
 & \leq \sum_{j \geq j_0} \mathbb{P} \left( y_j \geq \frac{1}{6M} \right) \\
 & \leq 6M \sum_{j \geq j_0} r^j \xrightarrow{j_0 \rightarrow +\infty} 0.
 \end{aligned}$$

On the other hand, assuming  $j_0 \geq 3$ , it is easy to check that  $(y_j)_{j \geq j_0} \stackrel{d}{=} (Py_j)_{j \geq 2}$  where  $P := U_2 \dots U_{j_0-1}$  is a product of  $j_0 - 2$  independent  $\text{Beta}(\theta, 1)$  random variables, and we deduce

$$\begin{aligned}
 & \mathbb{P} \left( \exists k \in \mathbb{Z} \cap \left[ 1, \frac{A-M}{2M} \right], \exists j \geq j_0, \frac{k}{M+A} \leq y_j \leq \frac{k}{A-M} \right) \\
 & = \mathbb{P} \left( \exists k \in \mathbb{Z} \cap \left[ 1, \min \left( \frac{A-M}{2M}, P(M+A) \right) \right], \right. \\
 & \quad \left. \exists j \geq 2, \frac{k}{P(M+A)} \leq y_j \leq \frac{k}{P(A-M)} \right).
 \end{aligned}$$

Conditionally to  $P$ , the corresponding quantity is bounded by the probability that there is at least one point of  $\mathcal{X}$  located in the disjoint union

$$\bigcup_{k \in \mathbb{Z} \cap [1, \min(\frac{A-M}{2M}, P(M+A))]} \left[ \frac{k}{P(M+A)}, \frac{k}{P(A-M)} \right],$$

hence

$$\begin{aligned}
 & \mathbb{P} \left( \exists k \in \mathbb{Z} \cap \left[ 1, \frac{A-M}{2M} \right], \exists j \geq j_0, \frac{k}{M+A} \leq y_j \leq \frac{k}{A-M} \right) \\
 & \leq \mathbb{E} \left( 1 - \exp \left( - \sum_{1 \leq k \leq \min(\frac{A-M}{2M}, P(M+A))} \int_{\frac{k}{P(M+A)}}^{\frac{k}{P(A-M)}} \frac{\theta}{x} dx \right) \right) \\
 & \leq 1 - \mathbb{E} \left( \exp \left( -\theta \min \left( \frac{A-M}{2M}, P(M+A) \right) \log \left( \frac{A+M}{A-M} \right) \right) \right) \\
 & \leq 1 - \mathbb{E} \left( \exp \left( -\theta P \frac{2M(M+A)}{A-M} \right) \right) \\
 & \leq 1 - \mathbb{E} (\exp(-6M\theta P)).
 \end{aligned}$$

In addition,  $P$  converges almost surely to 0 when  $j_0$  goes to  $+\infty$ , and  $0 \leq \exp(-6M\theta P) \leq 1$ , then by dominated convergence  $\mathbb{E}(\exp(-6M\theta P)) \xrightarrow{j_0 \rightarrow +\infty} 1$ .

Consequently,

$$\sup_A \mathbb{P} \left( \exists j \geq j_0, \exists k \in \mathbb{Z} \setminus \{0\}, f \left( \frac{k}{y_j} - A \right) \neq 0 \right) \xrightarrow{j_0 \rightarrow \infty} 0. \quad (3.36)$$

Furthermore, it is easy to check that

$$\sum_{j=1}^{j_0} \sum_{k \in \mathbb{Z}} f \left( \frac{k - \Phi_j}{y_j} \right) = \sum_{(j,k) \in S_{j_0}} f \left( \frac{k - \Phi_j}{y_j} \right)$$

and

$$\sum_{j=1}^{j_0} \sum_{k \in \mathbb{Z} \setminus \{0\}} f \left( \frac{k}{y_j} - A \right) = \sum_{(j,k) \in S_{j_0}} f \left( \frac{k - \{Ay_j\}}{y_j} \right)$$

where  $S_{j_0} := \{(j, k) : 1 \leq j \leq j_0, |k| \leq M + 1\}$  is finite. Using the previous lemma and the continuous mapping theorem we deduce

$$\sum_{(j,k) \in S_{j_0}} f \left( \frac{k - \{Ay_j\}}{y_j} \right) \xrightarrow[A \rightarrow \infty]{d} \sum_{(j,k) \in S_{j_0}} f \left( \frac{k - \Phi_j}{y_j} \right). \quad (3.37)$$

Now, let  $g$  be a continuous and bounded function from  $\mathbb{R}$  to  $\mathbb{R}$ . For all  $j$  and  $A$ , let  $V_{j,A} := \sum_{k \in \mathbb{Z} \setminus \{0\}} f \left( \frac{k}{y_j} - A \right)$ , and  $V_j := \sum_{k \in \mathbb{Z}} f \left( \frac{k - \Phi_j}{y_j} \right)$ . For all  $j_0 \geq 1$ , denoting  $\Omega_{j_0} := \{\forall j > j_0, V_{j,A} = 0\}$ , we have

$$\begin{aligned} \mathbb{E} \left( g \left( \sum_{j \geq 1} V_{j,A} \right) \right) &= \mathbb{E} \left( g \left( \sum_{j \leq j_0} V_{j,A} \right) \mathbf{1}_{\Omega_{j_0}} \right) + \mathbb{E} \left( g \left( \sum_{j \geq 1} V_{j,A} \right) \mathbf{1}_{\Omega_{j_0}^c} \right) \\ &= \mathbb{E} \left( g \left( \sum_{j \leq j_0} V_{j,A} \right) \right) - \mathbb{E} \left( g \left( \sum_{j \leq j_0} V_{j,A} \right) \mathbf{1}_{\Omega_{j_0}^c} \right) \\ &\quad + \mathbb{E} \left( g \left( \sum_{j \geq 1} V_{j,A} \right) \mathbf{1}_{\Omega_{j_0}^c} \right) \\ &= \mathbb{E} \left( g \left( \sum_{j \leq j_0} V_{j,A} \right) \right) + \mathcal{O}(\mathbb{P}(\Omega_{j_0}^c)). \end{aligned}$$

Hence

$$\begin{aligned}
 \overline{\lim}_{A \rightarrow \infty} \mathbb{E} \left( g \left( \sum_{j \geq 1} V_{j,A} \right) \right) &= \overline{\lim}_{A \rightarrow \infty} \mathbb{E} \left( g \left( \sum_{j \leq j_0} V_{j,A} \right) \right) + \mathcal{O}(\sup_A \mathbb{P}(\exists j \geq j_0, V_{j,A} \neq 0)) \\
 &= \mathbb{E} \left( g \left( \sum_{j \leq j_0} V_j \right) \right) + \mathcal{O}(\sup_A \mathbb{P}(\exists j \geq j_0, V_{j,A} \neq 0)) \\
 &= \mathbb{E} \left( g \left( \sum_{j \geq 1} V_j \right) \right) + \mathcal{O}(\mathbb{P}(\exists j \geq j_0, V_j \neq 0)) \\
 &\quad + \sup_A \mathbb{P}(\exists j \geq j_0, V_{j,A} \neq 0) \\
 &\stackrel{j_0 \rightarrow \infty}{=} \mathbb{E} \left( g \left( \sum_{j \geq 1} V_j \right) \right) + o(1)
 \end{aligned}$$

where the second equality derives from the convergence in distribution (3.37), and the last equality follows from (3.35) and (3.36). This gives (3.34).

*Remark 3.6.2.* This result can be easily extended to simple functions (linear combination of indicator functions) with compact support. In particular we have the following corollary:

**Corollary 3.6.3.** *Let  $s, t \in \mathbb{R}_+$ . Using the notations of Sections 3.3 and 3.5, we have*

$$X(s, s+t) \xrightarrow{s \rightarrow +\infty} \tilde{X}(t).$$

*Proof.* It suffices to write

$$\begin{aligned}
 X(s, s+t) &= \tau_\infty((s, s+t]) = \sum_{\substack{k \in \mathbb{Z} \setminus \{0\} \\ j \geq 1}} \mathbb{1}_{\frac{k}{y_j} - s \in (0, t]}, \\
 \tilde{X}(t) &= \tilde{\tau}_\infty((0, t]) = \sum_{\substack{k \in \mathbb{Z} \\ j \geq 1}} \mathbb{1}_{\frac{k - \Phi_j}{y_j} \in (0, t]},
 \end{aligned}$$

and for all  $x \in \mathbb{R}$ , a.s.,  $\sum_{\substack{k \in \mathbb{Z} \\ j \geq 1}} \mathbb{1}_{\frac{k - \Phi_j}{y_j} = x} = 0$ , so the continuous mapping theorem applies with  $f = \mathbb{1}_{(0, t]}$  under a similar reasoning as in the previous proof.  $\square$

## Appendix

In this section we prove Lemmas 3.5.1, 3.5.2 and 3.5.3.



**Proof of Lemma 3.5.1**

Let  $n \in \mathbb{N}^*$ . A simple change of variables ( $t = nx$ ) gives

$$\int_0^1 \frac{\{nx\}}{x} dx = \sum_{k=0}^{n-1} \int_0^1 \frac{t}{t+k} dt = 1 + \sum_{k=1}^{n-1} \int_0^1 \frac{t}{t+k} dt,$$

with for all  $k \geq 1$  and  $t \in [0, 1]$ ,  $\frac{1}{k+1} \leq \frac{1}{t+k} \leq \frac{1}{k}$ , thus

$$1 + \frac{1}{2} \sum_{k=1}^{n-1} \frac{1}{k+1} \leq \int_0^1 \frac{\{nx\}}{x} dx \leq 1 + \frac{1}{2} \sum_{k=1}^{n-1} \frac{1}{k}$$

so that

$$\int_0^1 \frac{\{nx\}}{x} dx \underset{n \rightarrow \infty}{=} \frac{1}{2} \log n + \mathcal{O}(1).$$

The same change of variables leads to

$$\begin{aligned} \int_0^1 \{nx\} \log x dx &= \sum_{k=0}^{n-1} \int_0^1 \frac{t}{n} \log \left( \frac{t+k}{n} \right) dt \\ &= -\frac{1}{2} \log n + \frac{1}{n} \sum_{k=0}^{n-1} \int_0^1 t \log(t+k) dt \\ &= -\frac{1}{2} \log n + \frac{1}{n} \sum_{k=0}^{n-1} \left( \frac{1}{2} \log(1+k) - \frac{1}{2} \int_0^1 \frac{t^2}{t+k} dt \right) \\ &= -\frac{1}{2} \log n + \frac{1}{2n} \log(n!) - \frac{1}{4n} - \frac{1}{2n} \sum_{k=1}^{n-1} \left( \frac{1}{2} - k + k^2 \log \left( 1 + \frac{1}{k} \right) \right). \end{aligned}$$

Moreover, as a consequence of Stirling's formula,

$$\log(n!) \underset{n \rightarrow \infty}{=} n \log n - n + \frac{1}{2} \log n + \mathcal{O}(1),$$

and furthermore we have

$$\frac{1}{2} - k + k^2 \log \left( 1 + \frac{1}{k} \right) \underset{k \rightarrow \infty}{=} \frac{1}{3k} + \mathcal{O} \left( \frac{1}{k^2} \right).$$

We deduce

$$\begin{aligned} \int_0^1 \{nx\} \log x dx &\underset{n \rightarrow \infty}{=} -\frac{1}{2} + \frac{1}{4n} \log n + \mathcal{O} \left( \frac{1}{n} \right) - \frac{1}{6n} \log n + \mathcal{O} \left( \frac{1}{n} \right) \\ &= -\frac{1}{2} + \frac{1}{12n} \log n + \mathcal{O} \left( \frac{1}{n} \right). \end{aligned}$$

**Proof of Lemma 3.5.2**

Let  $\ell, n \in \mathbb{N}^*$ . A summation by parts gives

$$\begin{aligned}
 \sum_{k=1}^n \frac{k}{n} \log \frac{k}{n} &= - \sum_{k=1}^{n-1} \log \left( 1 + \frac{1}{k} \right) \sum_{j=1}^k \frac{j}{n} \\
 &= - \sum_{k=1}^{n-1} \left( \frac{1}{k} + \mathcal{O} \left( \frac{1}{k^2} \right) \right) \frac{k(k+1)}{2n} \\
 &= - \sum_{k=1}^{n-1} \left( \frac{k}{2n} + \mathcal{O} \left( \frac{1}{n} \right) \right) \\
 &= - \frac{n(n+1)}{4n} + \mathcal{O}(1) \\
 &= - \frac{n}{4} + \mathcal{O}(1).
 \end{aligned}$$

Besides, for all fixed  $t \in (0, 1)$ ,

$$\begin{aligned}
 \sum_{k=1}^n \mathbb{1}_{\frac{k}{n} \geq t} \log \frac{k}{n} &= - \sum_{k=\lceil nt \rceil}^{n-1} \log \left( 1 + \frac{1}{k} \right) \sum_{j=\lceil nt \rceil}^k 1 \\
 &= - \sum_{k=\lceil nt \rceil}^{n-1} \left( \frac{1}{k} + \mathcal{O} \left( \frac{1}{k^2} \right) \right) (k - \lceil nt \rceil + 1) \\
 &= - \sum_{k=\lceil nt \rceil}^{n-1} \left( 1 - \frac{\lceil nt \rceil}{k} + \mathcal{O} \left( \frac{1}{k} \right) \right) \\
 &\stackrel{n \rightarrow \infty}{=} -n(1 - t + t \log t) + \mathcal{O}(1).
 \end{aligned}$$

Thus, on the one hand,

$$\begin{aligned}
 \sum_{k=1}^{n-1} \left\{ \ell \frac{k}{n} \right\} \log \frac{k}{n} &= \ell \left( \sum_{k=1}^{n-1} \frac{k}{n} \log \frac{k}{n} \right) - \sum_{m=1}^{\ell-1} \sum_{k=1}^{n-1} \mathbb{1}_{\frac{k}{n} \geq \frac{m}{\ell}} \log \frac{k}{n} \\
 &= \ell \left( -\frac{n}{4} + \mathcal{O}(1) \right) - \sum_{m=1}^{\ell-1} \left( -n \left( 1 - \frac{m}{\ell} + \frac{m}{\ell} \log \frac{m}{\ell} \right) + \mathcal{O}(1) \right) \\
 &= \left[ -\frac{\ell}{4} + \sum_{m=1}^{\ell-1} \left( 1 - \frac{m}{\ell} + \frac{m}{\ell} \log \frac{m}{\ell} \right) \right] n + \mathcal{O}(1) \\
 &= \left[ \frac{\ell}{4} - \frac{1}{2} + \sum_{m=1}^{\ell-1} \frac{m}{\ell} \log \frac{m}{\ell} \right] n + \mathcal{O}(1),
 \end{aligned}$$

and on the other hand,

$$\begin{aligned}
 \sum_{k=1}^{n-1} \left\{ -\ell \frac{k}{n} \right\} \log \frac{k}{n} &= -\ell \left( \sum_{k=1}^{n-1} \frac{k}{n} \log \frac{k}{n} \right) + \sum_{m=0}^{\ell-1} \sum_{k=1}^{n-1} \mathbb{1}_{\frac{k}{n} > \frac{m}{\ell}} \log \frac{k}{n} \\
 &= \left[ \frac{\ell}{4} - \sum_{m=1}^{\ell-1} \left( 1 - \frac{m}{\ell} + \frac{m}{\ell} \log \frac{m}{\ell} \right) \right] n + \mathcal{O}(1) + \sum_{k=1}^{n-1} \mathbb{1}_{\frac{k}{n} > 0} \log \frac{k}{n} \\
 &= \left[ -\frac{\ell}{4} + \frac{1}{2} - \sum_{m=1}^{\ell-1} \frac{m}{\ell} \log \frac{m}{\ell} \right] n + \mathcal{O}(1) + \sum_{k=1}^{n-1} \log \frac{k}{n}.
 \end{aligned}$$

Finally, it just remains to see

$$\sum_{k=1}^n \log \frac{k}{n} = - \sum_{k=1}^{n-1} k \log \left( 1 + \frac{1}{k} \right) = -n + \frac{1}{2} \log n + \mathcal{O}(1).$$

### Proof of Lemma 3.5.3

Let  $f : x \mapsto (\{px\} - \{qx\})^2$ . Denote for all positive integers  $m$ ,  $E_m = \{\frac{k}{m}; 1 \leq k \leq m-1\}$ , and let  $E_{p,q} = E_p \cup E_q$ . Noticing that  $x \mapsto (\{px\} - \{qx\})$  is a piecewise linear function with constant slope equal to  $p - q$  and jumps at multiples of  $1/p$  and multiples of  $1/q$ , the derivative of the distribution  $T_f$  related to  $f$  on  $(0, 1)$  is given by

$$\begin{aligned}
 (T_f)' &= T_{f'} + \sum_{r \in E_{p,q}} (f(r+0) - f(r-0)) \delta_r \\
 &= 2(p-q)(\{p\cdot\} - \{q\cdot\}) + \sum_{r \in E_{p,q}} (f(r+0) - f(r-0)) \delta_r
 \end{aligned}$$

Thus, integrating by parts gives

$$\int_0^1 \frac{f(x)}{x} dx = -2(p-q) \int_0^1 (\{px\} - \{qx\}) \log x dx - \sum_{r \in E_{p,q}} (f(r+0) - f(r-0)) \log r.$$

If  $r \in E_p \cap E_q$ , it is easy to check that  $f(r+0) - f(r-0) = 0$ .

If  $r = \frac{k}{p} \notin E_q$  then  $f(r+0) - f(r-0) = \left\{ q \frac{k}{p} \right\}^2 - \left( 1 - \left\{ q \frac{k}{p} \right\} \right)^2 = 2 \left\{ q \frac{k}{p} \right\} - 1$ .

Symmetrically if  $r = \frac{j}{q} \notin E_p$  then  $f(r+0) - f(r-0) = 2 \left\{ p \frac{j}{q} \right\} - 1$ .

Finally we get

$$\begin{aligned}
 & \sum_{r \in E_{p,q}} (f(r+0) - f(r-0)) \log r \\
 &= \sum_{\substack{1 \leq k \leq p-1 \\ \frac{1}{q} \nmid \frac{k}{p}}} \left( 2 \left\{ q \frac{k}{p} \right\} - 1 \right) \log \left( \frac{k}{p} \right) + \sum_{\substack{1 \leq j \leq q-1 \\ \frac{1}{p} \nmid \frac{j}{q}}} \left( 2 \left\{ p \frac{j}{q} \right\} - 1 \right) \log \left( \frac{j}{q} \right) \\
 &= \sum_{k=1}^{p-1} \left( 2 \left\{ q \frac{k}{p} \right\} - 1 \right) \log \left( \frac{k}{p} \right) + \sum_{j=1}^{q-1} \left( 2 \left\{ p \frac{j}{q} \right\} - 1 \right) \log \left( \frac{j}{q} \right) \\
 &\quad + 2 \sum_{m=1}^{\gcd(p,q)-1} \log \left( \frac{m}{\gcd(p,q)} \right).
 \end{aligned}$$

# 4. Characteristic polynomials at microscopic scale

This chapter is based on the article *Characteristic polynomials of modified permutation matrices at microscopic scale*, submitted in the journal Stochastic Processes and their Applications.

## Contents

---

4.1	Introduction . . . . .	<b>115</b>
4.1.1	Convergence of characteristic polynomials . . . . .	115
4.1.2	Generating random permutations . . . . .	116
4.1.3	Main results and outline of the chapter . . . . .	122
4.2	Proof of the main theorem . . . . .	<b>124</b>
4.2.1	Quotient of characteristic polynomials related to modified permutation matrices . . . . .	124
4.2.2	Quotient of characteristic polynomials related to permutation matrices . . . . .	128
4.3	Properties of the limiting function . . . . .	<b>134</b>
4.4	More general central measures . . . . .	<b>137</b>
4.4.1	Proof of Theorem 4.4.2 (i) . . . . .	139
4.4.2	Proof of Theorem 4.4.2 (ii) . . . . .	142
4.4.3	Properties of the limiting functions . . . . .	143

---

## 4.1 Introduction

### 4.1.1 Convergence of characteristic polynomials

Characteristic polynomials of random matrices have drawn much interest the last few decades. These objects encode the information of the whole spectrum of matrices. Moreover, in the case of unitarily invariant matrices (as Gaussian Unitary Ensemble or Circular Unitary Ensemble), the characteristic polynomial is believed to have a similar microscopic behavior as holomorphic functions which appear in number theory, as the Riemann zeta function. The characteristic polynomial of random matrices is also related to Gaussian fields, including the Gaussian multiplicative chaos introduced by Kahane [56].

On the macroscopic scale, Keating and Snaith [57], Hugues Keating and O-Connell [51], and then Bourgade Hugues Nikeghbali and Yor [24] study the logarithm of characteristic polynomial of unitary matrices following the Haar distribution, and prove in particular that its real and imaginary parts normalized by  $\sqrt{\frac{1}{2} \log n}$  converge jointly in law to independent centred and reduced Gaussian random variables.

Hambly, Keevash, O-Connell and Stark [48] give a similar result for permutation matrices following the uniform measure. Zeindler [92] [93] generalizes this result for permutation matrices under Ewens measures, considering more general class functions than the characteristic polynomial, the so-called multiplicative class functions. Dehaye and Zeindler [35], and Dang and Zeindler [34] extend the study to some Weyl groups, and some wreath products involving the symmetric group.

On the microscopic scale, Chhaibi, Najnudel and Nikeghbali [30] show that the characteristic polynomial of unitary matrices following the Haar measure, suitably renormalized, converges to a limiting entire function. With the coupling of virtual isometries introduced by Bourgade, Najnudel and Nikeghbali [25], the authors get an almost sure convergence. Chhaibi, Hovhannisyan, Najnudel, Nikeghbali, and Rodgers [28] extend the study to the special orthogonal group, the symplectic group, and give a related result for the Gaussian Unitary Ensemble.

Our motivation in this chapter is to prove similar results on the characteristic polynomial of some particular unitary matrices related to random permutations. More precisely:

- We focus on matrices belonging to two particular subgroups of the unitary group: the set of permutation matrices, and the wreath product  $S^1 \wr \mathfrak{S}_n$  (which can be seen as the set of permutation matrices where entries equals to one are replaced by complex numbers of modulus one).
- We tackle a large family of measures on the symmetric group, which are invariant by conjugation and verify a certain property of decay over the cycle lengths. This family includes the family of Ewens measures, as we shall see.
- We introduce a coupling method for generating sequences of modified permutations under these particular measures, by analogy of the notion of virtual isometries introduced in [25]. This coupling provides an almost sure convergence in our main result given below.

### 4.1.2 Generating random permutations

Before giving the construction of the random permutations we will deal with, let us recall the few following definitions and facts:

A **virtual permutation** is a sequence  $(\sigma_n)_{n \geq 1}$  where for each  $n$ ,  $\sigma_n$  is an element of  $\mathfrak{S}_n$  which can be derived from  $\sigma_{n+1}$  by simply removing the element  $n+1$  from the decomposition into disjoint cycles of  $\sigma_{n+1}$ . Let  $\mathfrak{S}$  denote the space of virtual permutations. We call **central measure** on  $\mathfrak{S}$  a probability measure which satisfies the following property:

$$\forall n \geq 1, \forall \tau \in \mathfrak{S}_n, \sigma_n \stackrel{d}{=} \tau \sigma_n \tau^{-1}.$$

For each  $n$ , it is easy to notice that every central measure on  $\mathfrak{S}_n$  can be fully described by a distribution on the set

$$\nabla^{(n)} := \left\{ (\ell_1, \dots, \ell_n) \in \mathbb{N}^n : \ell_1 \geq \ell_2 \geq \dots \geq \ell_n, \sum_{j=1}^n \ell_j = n \right\}$$

of partitions of the integer  $n$ , and conversely, in such a way that there is a one-to-one correspondence. A highly less obvious result (Theorem 2.3 in [74]) is that there exists a natural one-to-one correspondence between the central measures on  $\mathfrak{S}$  and the probability measures on

$$\nabla := \left\{ (x_1, x_2, \dots) \in [0, 1]^\infty : x_1 \geq x_2 \geq \dots, \sum_{j=1}^{+\infty} x_j \leq 1 \right\}. \quad (4.1)$$

The following definition introduces a new notion which specifies the family of measures we are going to consider in this chapter.

**Definition 4.1.1.** Let  $p$  be a probability measure on  $\nabla$ .

- We say that  $p$  is a **measure with exponential decay** if it satisfies the following property: There exists  $r \in (0, 1)$  and  $\nabla_1 \subset \nabla$  with  $p(\nabla_1) = 1$ , such that for all  $x = (x_1, x_2, \dots) \in \nabla_1$ ,

$$\exists C > 0, \forall j \geq 1, x_j \leq Cr^j. \quad (4.2)$$

- We say that a distribution on  $\mathfrak{S}$  is a **central measure with exponential decay** if its corresponding distribution on  $\nabla$  is a measure with exponential decay.

*Example 4.1.2.* The **Ewens measure** [40] of any arbitrary parameter  $\theta > 0$  on  $\mathfrak{S}$ , denoted by  $\text{Ewens}(\theta)$ , is a central measure with exponential decay.

Indeed, first recall that, given  $\theta > 0$ , one can define  $\text{Ewens}(\theta)$  on  $\mathfrak{S}$  thanks to the family of Ewens measures of parameter  $\theta$  on  $\mathfrak{S}_n$ ,  $n \geq 1$ , denoted by  $\text{Ewens}(n, \theta)$ , and defined by the probability functions

$$\forall \sigma \in \mathfrak{S}_n, \mathbb{P}_\theta^{(n)}(\sigma) = \frac{\theta^{K(\sigma)}}{\theta(\theta+1) \cdots (\theta+n-1)}, \quad (4.3)$$

where  $K(\sigma)$  denotes the total number of cycles of  $\sigma$  once decomposed into disjoint cycles. More precisely, the sequence of measures  $(\text{Ewens}(n, \theta))_{n \geq 1}$  is coherent with the projections  $\mathfrak{S}_{n+1} \rightarrow \mathfrak{S}_n$ . In other words, if  $\sigma_{n+1}$  follows  $\text{Ewens}(n+1, \theta)$ , then the random permutation obtained by removing the element  $n+1$  from the cycle-decomposition of  $\sigma_{n+1}$  follows  $\text{Ewens}(n, \theta)$ .

For each  $\theta > 0$ , the fact that  $\text{Ewens}(\theta)$  is central on  $\mathfrak{S}$  immediately derives from the fact that  $\text{Ewens}(n, \theta)$  is central on  $\mathfrak{S}_n$  for all  $n$ . It is also well-known that the corresponding distribution on  $\nabla$  of the central measure  $\text{Ewens}(\theta)$  is the **Poisson-Dirichlet distribution** of parameter  $\theta$  (denoted by  $\text{PD}(\theta)$ ).

Let  $y = (y_1, y_2, \dots)$  be a random vector following  $\text{PD}(\theta)$ . We know that  $Y$  has the same

distribution as the order statistics  $(Y_{(1)}, Y_{(2)}, \dots)$  of the random vector  $Y = (Y_1, Y_2, \dots)$  defined as follows: let  $(V_k)_{k \geq 1}$  be a sequence of i.i.d Beta(1,  $\theta$ ) random variables (with density function given by  $x \mapsto \theta(1-x)^{\theta-1} \mathbf{1}_{(0,1)}(x)$ ). For all  $j \geq 2$ , define  $Y_j := V_j \prod_{k=1}^{j-1} (1 - V_k)$ , and  $Y_1 := V_1$ . The distribution of  $Y$  is called GEM( $\theta$ ). In the literature, this method for generating such a vector  $Y$  with i.i.d random variables  $(V_k)$  is called **residual allocation model** [75] or **stick-breaking process** [59]. With this representation it is easy to compute that for all  $j$ ,

$$\mathbb{E}(Y_j) = \frac{1}{1+\theta} \left( \frac{\theta}{1+\theta} \right)^{j-1} \leq r_\theta^j$$

with  $r_\theta := \frac{1 \vee \theta}{1+\theta} < 1$ . Hence for any arbitrary  $r \in (r_\theta, 1)$ ,

$$\mathbb{P}(Y_j > r^j) \leq \frac{\mathbb{E}(Y_j)}{r^j} \leq \left( \frac{r_\theta}{r} \right)^j,$$

which is summable over  $j$ , and then the Borel-Cantelli lemma applies and gives that the number of  $j$  such that  $Y_j > r^j$  is almost surely finite. In other words, there exists a random number  $C > 0$  such that for all  $j$ ,  $Y_j \leq Cr^j$ . Finally, coming back to  $y$  it remains to see that the same kind of inequality holds for its coordinates, which is a direct consequence of the fact that for all  $j$  we have  $Y_{(j)} \leq ((Cr^k)_{k \geq 1})_{(j)} = Cr^j$ . Then the Ewens measure is a central measure with exponential decay.

*Remark 4.1.3.* Note that the Ewens measures are particular central measures whom corresponding distributions on  $\nabla$  are supported on

$$\nabla' := \left\{ (x_1, x_2, \dots) \in [0, 1]^\infty : x_1 \geq x_2 \geq \dots, \sum_{j=1}^{+\infty} x_j = 1 \right\} \subset \nabla. \quad (4.4)$$

In the main body of the chapter, we focus on central measures with exponential decay on  $\mathfrak{S}$  whose corresponding distributions on  $\nabla$  are supported on  $\nabla'$ .

Now, let us present the coupling we consider for generating random permutations, which is highly inspired from [85], [68], and [69].

Let  $\lambda = (\lambda_j)_{j \geq 1}$  be a sequence of decreasing real numbers summing to 1, and let  $E_\lambda = \bigsqcup_{j=1}^{\infty} \mathcal{C}_j$  be the disjoint union of circles  $\mathcal{C}_j$ , where for all  $j$ ,  $\mathcal{C}_j$  has perimeter  $\lambda_j$ . Let  $x = (x_k)_{k \geq 1} \in (E_\lambda)^\infty$ .

The coupling we are going to introduce is based on the following proposition:

**Proposition 4.1.4.** *One can define a virtual permutation  $\sigma_\infty(\lambda, x) = (\sigma_n(\lambda, x))_{n \geq 1}$ , where for all  $n$ , for all  $k \in \{1, \dots, n\}$ , the image of  $k$  by  $\sigma_n(\lambda, x)$  is the index of the first encountered point in  $\{x_1, \dots, x_n\}$  after  $x_k$  when exploring its circle counterclockwise. Moreover, if  $\lambda$  follows any arbitrary distribution  $p$  on  $\nabla'$ , and if conditionally on  $\lambda$  the points  $x_k$  are i.i.d following the uniform distribution on  $E_\lambda$ , then  $\sigma_\infty(\lambda, x)$  follows the central measure on  $\mathfrak{S}$  corresponding to  $p$ .*



*Example 4.1.5.* If  $\lambda$  follows the  $\text{PD}(\theta)$  distribution and if conditionally on  $\lambda$  the points  $x_k$  are i.i.d random variables uniformly distributed on  $E_\lambda$ , then  $\sigma_\infty(\lambda, x)$  follows  $\text{Ewens}(\theta)$ .

Let  $p$  be a distribution on  $\nabla'$ . Let  $y = (y_1, y_2, \dots)$  be a random vector following  $p$  and let  $E_y$  be the disjoint union of circles  $\mathcal{C}_j$  of perimeters  $y_j$ . Assume that conditionally given  $y$ , the  $x_k$  are i.i.d random variables uniformly distributed on  $E_y$ . Finally, introduce the array of random variables  $(\ell_{n,j})_{n,j \geq 1}$  defined by

$$\ell_{n,j} := \#\{k \in \{1, \dots, n\} : x_k \in \mathcal{C}_j\}, \quad (4.5)$$

and denote  $y_j^{(n)} := \frac{\ell_{n,j}}{n}$ .

Then, as a consequence of Proposition 4.1.4, almost surely,  $(y_1^{(n)}, y_2^{(n)}, \dots)$  converges in distribution to  $y$ . Moreover, conditionally on  $y$ , for all  $j$ ,

$$y_j^{(n)} = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{x_k \in \mathcal{C}_j} \xrightarrow[n \rightarrow \infty]{a.s.} y_j \quad (4.6)$$

by the strong law of large numbers.

In this chapter we also consider some modifications of permutation matrices, that we will call **modified permutation matrices**, which are permutation matrices where the entries equal to one are replaced by complex numbers of modulus one. The set of modified permutation matrices of size  $n$  has a group structure and can be identified to the wreath product  $S^1 \wr \mathfrak{S}_n$ , where  $S^1$  denotes the unit circle. Let us denote by  $\mathcal{T}_n$  the subset of matrices of  $S^1 \wr \mathfrak{S}_n$  which do not have 1 as an eigenvalue. The next lemma provides a construction of sequences of elements of  $\mathcal{T}_n$ ,  $n \geq 1$ , by analogy to the notion of **virtual isometries** introduced by Bourgade, Najnudel and Nikeghbali in [25].

**Lemma 4.1.6.** *For all  $n \geq 1$ , for all  $M \in \mathcal{T}_{n+1}$ , there exists a unique  $N \in \mathcal{T}_n$  such that*

$$\text{rank} \left( M - \left( \begin{array}{ccc|c} & & & 0 \\ & N & & \vdots \\ & & & 0 \\ \hline 0 & \dots & 0 & 1 \end{array} \right) \right) = 1. \quad (4.7)$$

*Moreover, the permutation corresponding to  $N$  derived from the one of  $M$  by removing the element  $n+1$  from its cycle-decomposition.*

*Proof.* Let  $n \geq 1$  and  $M \in \mathcal{T}_{n+1}$ . Write  $(w_1 \ w_2 \ \dots \ w_\ell \ w_{\ell+1} = n+1)$  the cycle of the corresponding permutation of  $M$  containing the element  $n+1$ . There exist  $z_1, \dots, z_\ell$  and  $z_{n+1}$  some complex numbers of modulus one such that for all  $k \in \{1, \dots, \ell\}$ ,  $Me_{w_k} = z_k e_{w_{k+1}}$  and  $Me_{n+1} = z_{n+1} e_{w_1}$  where  $(e_1, \dots, e_{n+1})$  is the canonical basis of  $\mathbb{C}^{n+1}$ .

Denote by  $M^{[n]}$  the top-left minor of size  $n$  of  $M$ . Let  $N \in \mathcal{M}_n(\mathbb{C})$ .

- If  $\ell = 0$  (i.e  $n+1$  is a fixed point of the associated permutation), then  $z_{n+1}$  is an eigenvalue of  $M$ . By hypothesis, this implies  $z_{n+1} \neq 1$ . Hence  $\text{rank}(M - \text{diag}(N, 1)) = 1$  if and only if  $N = M^{[n]}$  (since  $z_{n+1} - 1$  is the only non-zero entry of the last row and last column of  $M$ ). Moreover in this case, as  $M = \text{diag}(N, z_{n+1})$  we have  $N \in \mathcal{T}_n$ , and the procedure amounts to remove the fixed point  $n+1$  from the associated permutation of  $M$ .

- If  $\ell \geq 1$ , then  $Me_{w_\ell} = z_\ell e_{n+1}$  and  $Me_{n+1} = z_{n+1}e_{w_1}$  with  $w_1 \neq n+1 \neq w_\ell$ . The  $(\ell+1)$ -th roots of  $z_1 \dots z_\ell z_{n+1}$  are eigenvalues of  $M$ . By hypothesis, it follows  $z_1 \dots z_\ell z_{n+1} \neq 1$ . Moreover,  $\text{rank}(M - \text{diag}(N, 1)) = 1$  if and only if  $N = M^{[n]} + z_\ell z_{n+1} E_{w_1 w_\ell}$  where  $E_{ij}$  is the  $n$ -by- $n$  matrix with 1 in row  $i$  column  $j$ , and zeros elsewhere. In this case,  $Ne_{w_k} = z_k e_{w_{k+1}}$  for all  $k \in \{1, \dots, \ell-1\}$  (not considered when  $\ell = 1$ ) and  $Ne_{w_\ell} = z_\ell z_{n+1} e_{w_1}$ , so that the  $\ell$ -th roots of  $z_1 \dots z_\ell z_{n+1}$  are eigenvalues of  $N$  (the corresponding cycle is  $(w_1 \ w_2 \ \dots \ w_\ell)$ ). As  $z_1 \dots z_\ell z_{n+1} \neq 1$  we deduce  $N \in \mathcal{T}_n$ .

□

**Definition 4.1.7.** We say that a sequence of matrices  $(\widetilde{M}_n)_{n \geq 1}$  is a **modified virtual permutation** if for all  $n$ ,  $\widetilde{M}_n \in \mathcal{T}_n$  and  $\text{rank}(\widetilde{M}_{n+1} - \text{diag}(\widetilde{M}_n, 1)) = 1$ .

*Remark 4.1.8.* Note that every modified virtual permutation is in particular a virtual isometry.

**Proposition 4.1.9.** Let  $(\widetilde{M}_n)_{n \geq 1}$  be a modified virtual permutation. There exists a virtual permutation  $(\sigma_n)_{n \geq 1}$  such that, for all  $n \geq 1$ ,  $\widetilde{M}_n$  corresponds to the permutation  $\sigma_n$  and has a characteristic polynomial of the form

$$\chi_{\widetilde{M}_n}(X) := \det(XI - \widetilde{M}_n) = \prod_{\substack{j \geq 1 \\ \ell_{n,j} > 0}} (X^{\ell_{n,j}} - u_j),$$

where  $(u_j)_{j \geq 1}$  is a sequence of elements of  $S^1 \setminus \{1\}$  and the  $\ell_{n,j}$  denote the cycle-lengths of  $\sigma_n$ . Moreover, for all  $j$  and  $n$  such that  $\ell_{n,j} > 0$ ,  $u_j$  can be defined as the product of the non-zero entries of  $\widetilde{M}_n$  corresponding to the cycle  $j$  of  $\sigma_n$ .

*Proof.* Let  $n \geq 1$ . As in the proof of the previous lemma, let us denote by  $(w_1 \ w_2 \ \dots \ w_\ell \ w_{\ell+1} = n+1)$  the cycle of  $\sigma_{n+1}$  containing the element  $n+1$ , and by  $z_1, \dots, z_\ell$  and  $z_{n+1}$  the complex numbers of modulus one such that for all  $k \in \{1, \dots, \ell\}$ ,  $\widetilde{M}_{n+1}e_{w_k} = z_k e_{w_{k+1}}$  and  $\widetilde{M}_{n+1}e_{n+1} = z_{n+1}e_{w_1}$  where  $(e_1, \dots, e_{n+1})$  is the canonical basis of  $\mathbb{C}^{n+1}$ . The characteristic polynomials of  $\widetilde{M}_{n+1}$  and  $\widetilde{M}_n$  satisfy the equality

$$\chi_{\widetilde{M}_{n+1}}(X) = \begin{cases} \frac{X^{\ell+1} - z_1 \dots z_\ell z_{n+1}}{X^\ell - z_1 \dots z_\ell z_{n+1}} \chi_{\widetilde{M}_n}(X) & \text{if } \ell \geq 1 \\ (X - z_{n+1}) \chi_{\widetilde{M}_n}(X) & \text{if } \ell = 0 \end{cases}. \quad (4.8)$$

Indeed, if  $\ell = 0$ , then  $\widetilde{M}_{n+1}$  can be written  $\widetilde{M}_{n+1} = \text{diag}(\widetilde{M}^{[n]}, z_{n+1}) = \text{diag}(\widetilde{M}_n, z_{n+1})$  by the previous lemma, so that  $\chi_{\widetilde{M}_{n+1}}(X) = (X - z_{n+1}) \chi_{\widetilde{M}_n}(X)$ .

Otherwise, there exists a permutation matrix  $P$  of size  $n+1$  which fixes the element  $n+1$ , such that  $P\widetilde{M}_{n+1}P^{-1}$  and  $P\text{diag}(\widetilde{M}_n, 1)P^{-1}$  are block diagonal matrices where:

- All the blocks are of the form  $(\alpha_1)$  or  $\begin{pmatrix} 0 & \dots & 0 & \alpha_j \\ \alpha_1 & \ddots & & 0 \\ & \ddots & \ddots & \vdots \\ (0) & & \alpha_{j-1} & 0 \end{pmatrix}$ , with  $\alpha_1, \dots, \alpha_j \in S^1$ .

- If  $k$  is the number of blocks of  $P\widetilde{M}_{n+1}P^{-1}$ , then  $P\text{diag}(\widetilde{M}_n, 1)P^{-1}$  has exactly  $k+1$  blocks (including the bottom-right 1), and the  $k-1$  first blocks of  $P\widetilde{M}_{n+1}P^{-1}$  and  $P\text{diag}(\widetilde{M}_n, 1)P^{-1}$  are equal.

The last block of  $P\widetilde{M}_{n+1}P^{-1}$  is  $\begin{pmatrix} 0 & \dots & 0 & z_{n+1} \\ z_1 & \ddots & & 0 \\ & \ddots & \ddots & \vdots \\ (0) & & z_\ell & 0 \end{pmatrix}$ , hence with the help of the

previous lemma the penultimate block of  $P\text{diag}(\widetilde{M}_n, 1)P^{-1}$  is  $\begin{pmatrix} 0 & \dots & 0 & z_\ell z_{n+1} \\ z_1 & \ddots & & 0 \\ & \ddots & \ddots & \vdots \\ (0) & & z_{\ell-1} & 0 \end{pmatrix}$ ,

and we get

$$\begin{aligned} \chi_{\widetilde{M}_n}(X) &= \chi_{P^{[n]}\widetilde{M}_n(P^{-1})^{[n]}}(X) = \frac{\chi_{P\text{diag}(\widetilde{M}_n, 1)P^{-1}}(X)}{X-1} \\ &= \frac{\chi_{P\widetilde{M}_{n+1}P^{-1}}(X)}{X^{\ell+1} - z_1 \dots z_\ell z_{n+1}} (X^\ell - z_1 \dots z_{\ell-1} (z_\ell z_{n+1})) \end{aligned}$$

which gives (4.8). As a consequence, by analogy with the Chinese restaurant process (see *e.g.* [76]), here the customers arrive one by one and choose a table according to its weight, regardless of the past, and when a new customer  $n+1$  seats at a table (empty or not), it does not affect the element of  $S^1 \setminus \{1\}$  corresponding to this table. Hence we can assign a  $u_j$  to each table  $j$ , independently of  $n$ .  $\square$

**Definition 4.1.10.** We call **random modified permutation matrix** a random matrix  $\widetilde{M}_n$  such that:

- $\widetilde{M}_n$  corresponds to a random permutation  $\sigma_n$  generated with the procedure of Proposition 4.1.4 for a given distribution  $p$  on  $\nabla'$ .
- The non-zero entries of  $\widetilde{M}_n$  are i.i.d random variables uniformly distributed on the unit circle.

**Corollary 4.1.11.** Let  $(\sigma_n)_{n \geq 1}$  be a random virtual permutation, and let  $(u_j)_{j \geq 1}$  be a sequence of i.i.d uniform variables on the unit circle, independent of  $(\sigma_n)_{n \geq 1}$ . One can couple  $((\sigma_n)_{n \geq 1}, (u_j)_{j \geq 1})$  with a random modified virtual permutation  $(\widetilde{M}_n)_{n \geq 1}$  such that, for all  $n \geq 1$ ,

- $\widetilde{M}_n$  is a random modified permutation matrix corresponding to  $\sigma_n$ .
- Denoting by  $\ell_{n,j}$  the cycle-lengths of  $\sigma_n$ , then for all  $j$  and  $n$  such that  $\ell_{n,j} > 0$ ,  $u_j$  is the product of the non-zero entries of  $\widetilde{M}_n$  corresponding to the cycle  $j$  of  $\sigma_n$ .

*Proof.* This immediately derives from Proposition 4.1.9 and the fact that the projection  $M \mapsto N$  via  $\text{rank}(M - \text{diag}(N, 1))$  is coherent with respect to the sequence of probability measures  $(\mathcal{L}_n)$  defined for all  $n$  as the law of a  $n$ -by- $n$  random modified permutation matrix. Indeed, if the non-zero entries of  $\widetilde{M}_{n+1}$ , say  $z_1, z_2, \dots, z_{n+1}$ , are i.i.d uniform

variables on the unit circle, then the non-zero entries of  $\widetilde{M}_n$ , say  $z'_1, z'_2, \dots, z'_n$ , satisfy the following rule: There exists  $\pi \in \mathfrak{S}_{n+1}$  such that for all  $j \in \{1, \dots, n-1\}$ ,  $z'_j = z_{\pi(j)}$ , and  $z'_n = z_{\pi(n)}z_{\pi(n+1)}$ . Consequently  $z'_1, z'_2, \dots, z'_n$  are i.i.d uniform variables on the unit circle.  $\square$

### Notations

For all events  $A$  and all random variables  $y$ , we will denote by  $\mathbb{P}_y(A) := \mathbb{E}(\mathbb{1}_A \mid y)$  the conditional expectation of  $\mathbb{1}_A$  given  $y$ .

We will write  $X_n \xrightarrow[n \rightarrow \infty]{d} X$  for the convergence in distribution of the sequence of random variables  $(X_n)$  to the random variable  $X$ .

We will use the arrow  $\implies$  to denote the convergence in law on the space of continuous functions from  $\mathbb{C}$  to  $\mathbb{C}$  equipped with the topology of uniform convergence on compact sets.

Finally, for all real numbers  $x$ ,  $\{x\} = x - \lfloor x \rfloor$  will denote the fractional part of  $x$ , and  $\|x\|$  the distance from  $x$  to the nearest integer.

#### 4.1.3 Main results and outline of the chapter

Let  $\sigma = (\sigma_n)_{n \geq 1}$  be a random virtual permutation generated with the procedure given above. Let  $(M_n)_{n \geq 1}$  be the sequence of random permutation matrices associated to  $\sigma$ , that is to say for each  $n$  we define  $M_n$  as the  $n \times n$  matrix whose coordinates are given by

$$\forall 1 \leq i, j \leq n, (M_n)_{i,j} := \mathbb{1}_{i=\sigma_n(j)}. \quad (4.9)$$

Let  $(\widetilde{M}_n)_{n \geq 1}$  be a random modified virtual permutation generated by  $\sigma$  and a sequence  $(u_j)_{j \geq 1}$  of i.i.d uniform variables on the unit circle independent of  $\sigma$  (see Corollary 4.1.11). In particular, using the notations of eq. (4.5), for all  $n$  and  $j$  such that  $\ell_{n,j} > 0$ ,  $u_j$  is the product of the non-zero entries of  $\widetilde{M}_n$  whom cycle is associated with the circle  $\mathcal{C}_j$ , so that  $u_j$  does not depend on  $n$ .

For all  $n \in \mathbb{N}^*$  and  $z \in \mathbb{C}$ , we consider the characteristic polynomials of  $M_n$  and  $\widetilde{M}_n$ , respectively defined by

$$Z_n(z) := \det(zI - M_n) \quad (4.10)$$

and

$$\widetilde{Z}_n(z) := \det(zI - \widetilde{M}_n). \quad (4.11)$$

Let  $\alpha$  be an irrational number between 0 and 1, and set

$$\widetilde{\xi}_n(z) = \frac{\widetilde{Z}_n(e^{2i\pi z/n})}{\widetilde{Z}_n(1)} \quad (4.12)$$

and

$$\xi_{n,\alpha}(z) = \frac{Z_n\left(e^{2i\pi\left(\frac{z}{n} + \alpha\right)}\right)}{Z_n(e^{2i\pi\alpha})}. \quad (4.13)$$

The characteristic polynomial  $\widetilde{\xi}_n$  can be written with the help of  $(y_j^{(n)})_{j \geq 1}$  and  $(u_j)_{j \geq 1}$

as

$$\tilde{\xi}_n(z) = \prod_{\substack{j \geq 1 \\ \ell_{n,j} > 0}} \frac{e^{2i\pi z y_j^{(n)}} - u_j}{1 - u_j}. \quad (4.14)$$

The characteristic polynomial  $\xi_{n,\alpha}$  can be written as

$$\xi_{n,\alpha} = \prod_{\substack{j \geq 1 \\ \ell_{n,j} > 0}} \frac{e^{2i\pi(\frac{z}{n} + \alpha)\ell_{n,j}} - 1}{e^{2i\pi\alpha\ell_{n,j}} - 1}. \quad (4.15)$$

We will give more details about the expressions of  $\tilde{\xi}_n$  and  $\xi_{n,\alpha}$  in the next section.

Finally, let us recall that the type of any real number  $x$  is defined by

$$\eta = \sup\{\gamma \in \mathbb{R} : \liminf_{n \rightarrow +\infty} n^\gamma \|nx\| = 0\} \in \mathbb{R}_+ \cup \{+\infty\}. \quad (4.16)$$

We say that  $x$  is of **finite type** if  $\eta$  is finite. A basic property is that if  $x$  is irrational then its type is greater or equal to one (and can be infinite).

We are ready to state the main result of the chapter:

**Theorem 4.1.12.** *Assume that  $\sigma$  is generated by the coupling described above for a distribution  $p$  on  $\nabla'$  with exponential decay. Then we have the following convergences:*

- (i) *Almost surely,  $\tilde{\xi}_n$  converges uniformly on every compact set to an entire function  $\tilde{\xi}_\infty$  defined by*

$$\tilde{\xi}_\infty(z) = \prod_{j=1}^{+\infty} \frac{e^{2i\pi z y_j} - u_j}{1 - u_j}.$$

- (ii) *Assume  $\alpha$  is an irrational number of finite type. Then*

$$\xi_{n,\alpha} \xrightarrow[n \rightarrow \infty]{} \tilde{\xi}_\infty$$

where  $\tilde{\xi}_\infty$  is the same entire function as above.

*Remark 4.1.13.* Without the coupling, the theorem still holds replacing the first point by:

(i')

$$\tilde{\xi}_n \xrightarrow[n \rightarrow \infty]{} \tilde{\xi}_\infty.$$

*Remark 4.1.14.* Note that the parameter  $\alpha$  is not allowed to be rational, otherwise some denominators in the product expression of  $\xi_{n,\alpha}$  could be zeros. Moreover, heuristically, the motivation to take  $\alpha$  irrational of finite type is to avoid a too fast accumulation of small denominators.

The chapter is organized as follows: In Section 4.2, we give a proof of Theorem 4.1.12 by showing the first point in Subsection 4.2.1, and the second point in Subsection 4.2.2. These two subsections are mutually independent. In Section 4.3, we give some estimates on the limiting function  $\tilde{\xi}_\infty$ , and compare our results to the unitary case presented in

[30]. Finally, in Section 4.4, we extend the study to more general central measures, removing the restriction to  $\nabla'$  for the support of their corresponding distributions on  $\nabla$ .

## 4.2 Proof of the main theorem

### 4.2.1 Quotient of characteristic polynomials related to modified permutation matrices

Consider a distribution with exponential decay  $p$  on  $\nabla'$ , giving a  $r \in (0, 1)$  as in (4.2). Let  $y = (y_1, y_2, \dots)$  be a random vector following the distribution  $p$ .

Let  $(\widetilde{M}_n)_{n \geq 1}$  be a sequence of modified random permutation matrices generated by the coupling given by Corollary 4.1.11. For all  $n \in \mathbb{N}^*$  and  $z \in \mathbb{C}$ , we consider the characteristic polynomial of  $\widetilde{M}_n$ , defined by (4.11).

As one is almost surely not a zero of  $\widetilde{Z}_n$ , the function  $\widetilde{\xi}_n$  defined by (4.12) is an entire function.

Using Corollary 4.1.11,  $\widetilde{\xi}_n$  can be reformulated with the help of a sequence  $(u_j)_{j \geq 1}$  of independent random variables that are uniformly distributed on the unit circle (and independent of the  $y_k^{(n)}$ ) as

$$\begin{aligned} \widetilde{\xi}_n(z) &= \prod_{\substack{j \geq 1 \\ \ell_{n,j} > 0}} \frac{e^{2i\pi z y_j^{(n)}} - u_j}{1 - u_j} \\ &= \prod_{\substack{j \geq 1 \\ \ell_{n,j} > 0}} \left( 1 + \frac{1}{1 - u_j} (e^{2i\pi z y_j^{(n)}} - 1) \right). \end{aligned} \tag{4.17}$$

The next lemmas aim to handle the tail of the infinite product in the expression of  $\widetilde{\xi}_n(z)$ , in order to apply a dominated convergence theorem and get the pointwise convergence of  $\widetilde{\xi}_n$ . Moreover, they provide a bound of  $\widetilde{\xi}_n$  uniformly on compact sets, allowing to conclude with Montel theorem.

**Lemma 4.2.1.** *Let  $\alpha > 2$ . For all  $k \in \mathbb{N}^*$ , set  $m_k := \min_{1 \leq j \leq k} |1 - u_j|$ . Then a.s there exists a random number  $C_1 > 0$  such that for all  $k$ ,*

$$m_k > C_1 k^{-\alpha}. \tag{4.18}$$

*Proof.* Let  $A \in (0, 1)$ . Let  $T$  be a random variable following the uniform distribution on  $[0, 1]$ .

$$\mathbb{P}(|1 - e^{2i\pi T}| \geq A) \geq \mathbb{P}(\sin(\pi T) \geq A) \geq \mathbb{P}(2 \min(T, 1 - T) \geq A) = \mathbb{P}(T \geq A) = 1 - A.$$

Then for all  $k$ ,

$$\mathbb{P}(m_k \leq A) \leq 1 - (1 - A)^k \leq kA$$

using the mean value inequality. Thus,

$$\sum_{k=1}^{+\infty} \mathbb{P}(m_k \leq k^{-\alpha}) \leq \sum_{k=1}^{+\infty} k^{1-\alpha} < +\infty.$$

Applying Borel-Cantelli lemma we deduce that the number of  $k$  such that  $m_k \leq k^{-\alpha}$  is a.s finite, i.e a.s there exists  $k_0 \in \mathbb{N}^*$  such that for all  $k > k_0$ ,  $m_k > k^{-\alpha}$ . Finally,  $C_1 := \min_{j \leq k_0} (j^\alpha |1 - u_j|) \wedge 1$  gives the claim.  $\square$

**Lemma 4.2.2.** *For all  $\rho \in (\sqrt{r}, 1)$ , a.s there exists a random number  $C_2 > 0$ , such that for all  $j \geq 1$ ,*

$$s_j := \sup_{n \geq 1} y_j^{(n)} \leq C_2 \rho^j. \quad (4.19)$$

*Proof.* Let  $\rho \in (0, 1)$ . For each  $j$  and  $n$ , write  $y_j^{(n)}$  as the empirical mean of  $n$  i.i.d Bernoulli random variables  $Z_{1,j}, Z_{2,j}, \dots, Z_{n,j}$  of parameter  $y_j$ . Fix  $j$ .

If  $n \leq \rho^{-j}$ , then it is easy to check that the events  $\{y_j^{(n)} \geq \rho^j\}$  and  $\{\exists k \in \{1, 2, \dots, n\} : Z_{k,j} = 1\}$  are equal, hence

$$\mathbb{E} \left( \mathbb{1}_{y_j^{(n)} \geq \rho^j} \mid y_j \right) = \mathbb{E} \left( \mathbb{1}_{\exists k \in \{1, 2, \dots, n\} : Z_{k,j} = 1} \mid y_j \right) \leq n y_j \leq \frac{y_j}{\rho^j},$$

and then

$$\sum_{n \leq \rho^{-j}} \mathbb{P}_y(y_j^{(n)} \geq \rho^j) \leq \frac{y_j}{\rho^{2j}}.$$

If  $n \geq \rho^{-j}$ , then for any arbitrary  $\lambda > 0$  we have the Chernoff bound

$$\begin{aligned} \mathbb{E} \left( \mathbb{1}_{y_j^{(n)} \geq \rho^j} \mid y_j \right) &\leq e^{-\lambda \rho^j} \mathbb{E} \left[ e^{\frac{\lambda}{n} Z_{1,j}} \mid y_j \right]^n \\ &= e^{-\lambda \rho^j} \left( 1 - y_j + y_j e^{\frac{\lambda}{n}} \right)^n \\ &\leq e^{-\lambda \rho^j} \exp \left( n y_j \left( e^{\frac{\lambda}{n}} - 1 \right) \right). \end{aligned}$$

This inequality is optimized at point  $\lambda = n \log(\rho^j / y_j)$ , which gives

$$\mathbb{E} \left( \mathbb{1}_{y_j^{(n)} \geq \rho^j} \mid y_j \right) \leq e^{-n \left( \rho^j \log \left( \frac{\rho^j}{y_j} \right) - \rho^j + y_j \right)} \leq e^{-n \rho^j \left( \log \left( \frac{\rho^j}{y_j} \right) - 1 \right)}.$$

By assumption of exponential decay (4.2), conditionally on  $y = (y_1, y_2, \dots)$ , almost surely there exists a constant  $C > 0$  such that for all  $j \geq 1$ ,  $y_j \leq C r^j$ . Thus, taking any arbitrary  $r' \in (r, 1)$ , for almost every  $y$  there exists an integer  $k$  such that for all  $j \geq k$ ,  $y_j \leq r'^j$ . Fix a given  $y$ ,  $r'$  and  $k$ . Then, setting  $\rho > \sqrt{r'}$ ,

$$\log \left( \frac{\rho^j}{y_j} \right) \geq \log \left( \frac{\rho^j}{r'^j} \right) = j(\log \rho - \log r') \geq 2$$

for all  $j$  sufficiently large and greater than  $k$ , say for all  $j \geq m$  ( $m$  dependent on  $y$ ).

Then for all  $j \geq m$ ,

$$\sum_{n \geq \rho^{-j}} \mathbb{P}_y(y_j^{(n)} \geq \rho^j) \leq \frac{\frac{y_j}{\rho^j} e}{1 - e^{-\rho^j \left( \log\left(\frac{\rho^j}{y_j}\right) - 1 \right)}} \leq \frac{\frac{y_j}{\rho^j} e}{1 - e^{-\rho^j}} \leq (2e) \frac{y_j}{\rho^{2j}}.$$

We deduce, for all  $j \geq m$ ,

$$\sum_{n=1}^{+\infty} \mathbb{P}_y(y_j^{(n)} \geq \rho^j) \leq (1 + 2e) \frac{y_j}{\rho^{2j}} \leq (1 + 2e) \left( \frac{r'}{\rho^2} \right)^j$$

and consequently, we get

$$\mathbb{E} \left[ \sum_{j=m}^{+\infty} \sum_{n=1}^{+\infty} \mathbb{1}_{y_j^{(n)} \geq \rho^j} \mid y \right] < +\infty$$

which implies

$$\mathbb{P}_y \left( \sum_{j=m}^{+\infty} \sum_{n=1}^{+\infty} \mathbb{1}_{y_j^{(n)} \geq \rho^j} < +\infty \right) = 1$$

for almost every  $y$ . Finally, taking the expectation we get

$$\mathbb{P} \left( \sum_{j=m}^{+\infty} \sum_{n=1}^{+\infty} \mathbb{1}_{y_j^{(n)} \geq \rho^j} < +\infty \right) = 1.$$

In other words, the number of couples  $(j, n)$  such that  $j \geq m$  and  $y_j^{(n)} \geq \rho^j$  is almost surely finite, which gives that the number of  $j$  such that  $\sup_{n \in \mathbb{N}^*} y_j^{(n)} \geq \rho^j$  is almost surely finite.  $\square$

**Lemma 4.2.3.** *With the same notation as above, a.s.,*

$$C_3 := \sum_{j=1}^{+\infty} \frac{y_j}{|1 - u_j|} < +\infty \quad (4.20)$$

and

$$C_4 := \sum_{j=1}^{+\infty} \frac{s_j}{|1 - u_j|} < +\infty. \quad (4.21)$$

*Proof.* Straightforward consequence of Lemma 4.2.1 and Lemma 4.2.2.  $\square$

Now, we are able to prove the first point of Theorem 4.1.12:

*Proof of Theorem 4.1.12 (i).* We begin to show the pointwise convergence. Let  $z \in \mathbb{C}$ . The idea of the proof consists in splitting the product in the expression of  $\tilde{\xi}_n(z)$  as follows:

$$\tilde{\xi}_n(z) = \prod_{j=1}^{j_0} \frac{e^{2i\pi z y_j^{(n)}} - u_j}{1 - u_j} \prod_{j=j_0+1}^{+\infty} \frac{e^{2i\pi z y_j^{(n)}} - u_j}{1 - u_j}$$



where  $j_0$  is an integer depending on  $|z|$  and on random numbers  $C_1$  and  $C_2$ , chosen in such a way that for all  $j > j_0$ ,

$$\left| \frac{1}{1-u_j} (e^{2i\pi z y_j^{(n)}} - 1) \right| \leq \frac{9}{10} < 1.$$

Indeed we can chose such a  $j_0$  as for all  $j$ ,

$$\begin{aligned} \left| \frac{1}{1-u_j} (e^{2i\pi z y_j^{(n)}} - 1) \right| &\leq \frac{1}{C_1 j^{-3}} 2\pi |z| y_j^{(n)} \exp(2\pi |z| y_j^{(n)}) \\ &\leq \frac{1}{C_1 j^{-3}} 2\pi |z| C_2 \rho^j \exp(2\pi |z| C_2 \rho^j) \end{aligned}$$

and then it suffices to take  $j$  large enough such that  $\max(\frac{1}{C_1 j^{-3}} 2\pi |z| C_2 \rho^j, 2\pi |z| C_2 \rho^j) \leq \frac{1}{2}$ , which provides a bound lower than  $\frac{1}{2} e^{1/2} \approx 0.82 \leq \frac{9}{10}$ . Thus, we can apply the logarithm to the product of terms for  $j > j_0$  in the expression of  $\tilde{\xi}_n$ , and furthermore it is straightforward to check that for all  $Z \in \mathbb{C}$  such that  $|Z| \leq \frac{9}{10}$  we have

$$|\log(1+Z)| = \left| \sum_{k=1}^{+\infty} \frac{(-1)^{k+1}}{k} Z^k \right| \leq \sum_{k=1}^{+\infty} \frac{|Z|^k}{k} \leq \frac{|Z|}{1-|Z|} \leq 10|Z|.$$

Consequently for all  $j > j_0$ ,

$$\begin{aligned} \left| \log \left( 1 + \frac{1}{1-u_j} (e^{2i\pi z y_j^{(n)}} - 1) \right) \right| &\leq 10 \left| \frac{1}{1-u_j} (e^{2i\pi z y_j^{(n)}} - 1) \right| \\ &\leq \frac{10}{|1-u_j|} \sum_{k=1}^{+\infty} \frac{|2i\pi z y_j^{(n)}|^k}{k!} \\ &\leq \frac{10}{|1-u_j|} y_j^{(n)} (e^{2\pi |z|} - 1) \\ &\leq \frac{10 e^{2\pi |z|} C_2}{C_1} \rho^j j^3 \end{aligned}$$

which is summable in  $j$ . Moreover, as  $y_j^{(n)}$  converges a.s to  $y_j$ , then by continuity  $\log \left( 1 + \frac{1}{1-u_j} (e^{2i\pi z y_j^{(n)}} - 1) \right)$  converges to  $\log \left( 1 + \frac{1}{1-u_j} (e^{2i\pi z y_j} - 1) \right)$ . Hence by dominated convergence,

$$\tilde{\xi}_n(z) \xrightarrow{n \rightarrow \infty} \tilde{\xi}_\infty(z).$$

This holds true for every  $z \in \mathbb{C}$  so we get the pointwise convergence of  $\tilde{\xi}_n$  to  $\tilde{\xi}_\infty$  on  $\mathbb{C}$ . Now, consider an arbitrary compact set of  $\mathbb{C}$  included in  $\{z \in \mathbb{C} : |z| \leq K\}$  for a certain

fixed  $K > 0$ . For all  $z$  in this compact set and all  $n \geq 1$ ,

$$\begin{aligned}
 |\tilde{\xi}_n(z)| &\leq \prod_{j=1}^{+\infty} \left( 1 + \frac{1}{|1 - u_j|} |e^{2i\pi z y_j^{(n)}} - 1| \right) \\
 &\leq \prod_{j=1}^{+\infty} \left( 1 + \frac{y_j^{(n)}}{|1 - u_j|} (e^{2\pi K} - 1) \right) \\
 &\leq \exp \left( \sum_{j=1}^{+\infty} \frac{y_j^{(n)}}{|1 - u_j|} (e^{2\pi K} - 1) \right) \\
 &\leq \exp \left( (e^{2\pi K} - 1) \sum_{j=1}^{+\infty} \frac{s_j}{|1 - u_j|} \right) \\
 &= \exp (C_4 (e^{2\pi K} - 1)).
 \end{aligned}$$

We deduce by Montel theorem the uniform convergence of  $\tilde{\xi}_n$  to  $\tilde{\xi}_\infty$  on all compact sets.  $\square$

#### 4.2.2 Quotient of characteristic polynomials related to permutation matrices (without modification)

Here, the problem of finding a suitable normalization of the characteristic polynomial in order to have a non-trivial limiting function is more difficult than the previous one. We precise below the nature of this problem and defend our choice of normalization.

Consider a distribution with exponential decay  $p$  on  $\nabla'$ , giving a  $r \in (0, 1)$  as in (4.2), and let  $y = (y_1, y_2, \dots)$  be a random vector following the distribution  $p$ .

Let  $(M_n)_{n \geq 1}$  be a sequence of random permutation matrices generated by the coupling described in Subsection 4.1.2 with respect to  $y$ . For all  $n \in \mathbb{N}^*$  and  $z \in \mathbb{C}$ , we consider the characteristic polynomial of  $M_n$  defined by (4.10).

Contrarily to random permutation matrices with modification (the  $\tilde{M}_n$  defined in the previous subsection), for every  $n$ , there are some points  $z$  on the unit circle such that  $\mathbb{P}(Z_n(z) = 0) > 0$ . For instance, for all  $n$ , the characteristic polynomial of  $M_n$  evaluated at  $z = 1$  is almost surely zero (since each  $j$ -cycle of the associated permutation corresponds to eigenvalues which are exactly the  $j$ -th roots of unity). Thus, the function  $\tilde{\xi}_n$  of the previous section, replacing  $\tilde{Z}_n$  by  $Z_n$ , i.e

$$\tilde{\xi}_n(z) = \frac{Z_n(e^{2i\pi z/n})}{Z_n(1)}$$

is not well-defined on the whole complex plane here. Based on the fact that all eigenvalues of permutation matrices are roots of unity, then, for every irrational number  $\alpha$ ,  $z = e^{2i\pi\alpha}$  is almost surely not a zero of  $Z_n$  for all  $n$ . Let  $\alpha$  be an irrational number between 0 and 1. It is natural to shift the random process of eigenangles by  $2\pi\alpha$ , and consider the function  $\xi_{n,\alpha}$  defined by (4.13).

As  $\alpha$  is irrational, then  $\xi_{n,\alpha}$  is an entire function, and can be written as follows:

$$\begin{aligned}\xi_{n,\alpha} &= \prod_{\substack{j \geq 1 \\ \ell_{n,j} > 0}} \frac{e^{2i\pi(\frac{j}{n} + \alpha)\ell_{n,j}} - 1}{e^{2i\pi\alpha\ell_{n,j}} - 1} \\ &= \prod_{\substack{j \geq 1 \\ \ell_{n,j} > 0}} \left( 1 + \frac{e^{2i\pi\alpha\ell_{n,j}}}{e^{2i\pi\alpha\ell_{n,j}} - 1} \left( e^{2i\pi z y_j^{(n)}} - 1 \right) \right).\end{aligned}\tag{4.22}$$

Heuristically, the idea of considering  $\alpha$  of finite type in Theorem 4.1.12 is to get the denominators in the expression of  $\xi_{n,\alpha}$  not too close from 0 when  $n$  become large, by comparison to the factor  $e^{2i\pi z y_j^{(n)}} - 1$  in the numerator. In the sequel we prove the result into two parts: first, we show the convergence in distribution for finite products, and then, we handle the remaining infinite product.

**Lemma 4.2.4.** *For all  $k \in \mathbb{N}^*$ , conditionally on  $y$ ,*

$$(\ell_{n,1}, \ell_{n,2}, \dots, \ell_{n,k}, n - \ell_{n,1} - \ell_{n,2} - \dots - \ell_{n,k}) \stackrel{d}{=} \mathcal{M}(n, y_1, y_2, \dots, y_k, 1 - y_1 - y_2 - \dots - y_k),$$

where  $\mathcal{M}(n, q_1, q_2, \dots, q_m)$  denotes a multinomial random variable of parameters  $q_1, q_2, \dots, q_m$ .

*Proof.* Direct consequence of (4.5). □

**Lemma 4.2.5.** *For all  $k \in \mathbb{N}^*$ , conditionally on  $y$ ,*

$$\left( \{\alpha\ell_{n,1}\}, \{\alpha\ell_{n,2}\}, \dots, \{\alpha\ell_{n,k}\}, y_1^{(n)}, \dots, y_k^{(n)} \right) \xrightarrow[n \rightarrow \infty]{d} (\Phi_1, \Phi_2, \dots, \Phi_k, y_1, y_2, \dots, y_k)$$

where the  $\Phi_j$  are i.i.d random variables uniformly distributed on  $[0, 1]$ , independent of the  $y_j$ .

*Proof.* Let  $j_1, \dots, j_k$  be integers, and let  $\lambda_1, \dots, \lambda_k$  be real numbers. We have:

$$\begin{aligned}& \mathbb{E} \left( e^{2i\pi(j_1\{\alpha\ell_{n,1}\} + \dots + j_k\{\alpha\ell_{n,k}\}) + i(\lambda_1 \frac{\ell_{n,1}}{n} + \dots + \lambda_k \frac{\ell_{n,k}}{n})} \middle| y \right) \\ &= \mathbb{E} \left( e^{2i\pi(j_1\alpha + \frac{\lambda_1}{2\pi n})\ell_{n,1} + \dots + 2i\pi(j_k\alpha + \frac{\lambda_k}{2\pi n})\ell_{n,k} + 0 \times (n - \ell_{n,1} - \dots - \ell_{n,k})} \middle| y \right) \\ &= \sum_{\substack{(\ell_1, \dots, \ell_{k+1}) \in \mathbb{N}^{k+1} \\ \ell_1 + \dots + \ell_{k+1} = n}} \frac{n!}{\ell_1! \dots \ell_{k+1}!} \left( y_1 e^{2i\pi(j_1\alpha + \frac{\lambda_1}{2\pi n})} \right)^{\ell_1} \dots \left( y_k e^{2i\pi(j_k\alpha + \frac{\lambda_k}{2\pi n})} \right)^{\ell_k} \\ &\quad \times (1 - y_1 - \dots - y_k)^{\ell_{k+1}} \\ &= \left( y_1 e^{2i\pi(j_1\alpha + \frac{\lambda_1}{2\pi n})} + \dots + y_k e^{2i\pi(j_k\alpha + \frac{\lambda_k}{2\pi n})} + (1 - y_1 - \dots - y_k) \right)^n.\end{aligned}$$

If  $j_1 = j_2 = \dots = j_k = 0$ , then this quantity converges to  $e^{i(\lambda_1 y_1 + \dots + \lambda_k y_k)}$ .

Otherwise, there exists  $m \in \llbracket 1, k \rrbracket$  such that  $j_m \neq 0$ . Since  $\alpha$  is irrational,  $\|j_m \alpha\| \neq 0$ . Hence there exists  $N \in \mathbb{N}$  such that for all  $n \geq N$ ,  $|\frac{\lambda_m}{2\pi n}| \leq \frac{\|j_m \alpha\|}{2} < \frac{1}{2}$ . Then for all

$n \geq N$ ,

$$\left\| j_m \alpha + \frac{\lambda_m}{2\pi n} \right\| \geq \|j_m \alpha\| - \left\| \frac{\lambda_m}{2\pi n} \right\| \geq \frac{\|j_m \alpha\|}{2}.$$

Consequently,  $y_1 e^{2i\pi(j_1 \alpha + \frac{\lambda_1}{2\pi n})} + \dots y_k e^{2i\pi(j_k \alpha + \frac{\lambda_k}{2\pi n})} + (1 - y_1 - \dots - y_k)$  is a convex combination with a.s positive coefficients of points located on the unit circle, whose the distance between two points (namely the point 1 and the point  $e^{2i\pi(j_m \alpha + \frac{\lambda_m}{2\pi n})}$ ) is bounded from below by a positive number which is independent of  $n$ . Thus  $y_1 e^{2i\pi(j_1 \alpha + \frac{\lambda_1}{2\pi n})} + \dots y_k e^{2i\pi(j_k \alpha + \frac{\lambda_k}{2\pi n})} + (1 - y_1 - \dots - y_k)$  is bounded, uniformly in  $n$ , by a quantity strictly smaller than one, and finally the Fourier transform goes to 0, which gives the claim.  $\square$

**Proposition 4.2.6.** *For all  $k \in \mathbb{N}^*$ , conditionally on  $y$ ,*

$$\prod_{\substack{1 \leq j \leq k \\ \ell_{n,j} > 0}} \left( 1 + \frac{e^{2i\pi \alpha \ell_{n,j}}}{e^{2i\pi \alpha \ell_{n,j}} - 1} \left( e^{2i\pi z y_j^{(n)}} - 1 \right) \right) \xrightarrow{n \rightarrow \infty} \prod_{\substack{1 \leq j \leq k \\ \ell_{n,j} > 0}} \left( 1 + \frac{e^{2i\pi \Phi_j}}{e^{2i\pi \Phi_j} - 1} (e^{2i\pi z y_j} - 1) \right).$$

*Proof.* Let  $k \in \mathbb{N}^*$ . The map

$$(x_1, x_2, \dots, x_{2k}) \mapsto \left( z \mapsto \prod_{j=1}^k \left( 1 + \frac{e^{2i\pi x_j}}{e^{2i\pi x_j} - 1} (e^{2i\pi z x_{k+j}} - 1) \right) \right)$$

is defined and continuous on  $(\mathbb{R} \setminus \mathbb{Z})^k \times \mathbb{R}^k$ . Moreover,  $\mathbb{P}(\exists j \in \llbracket 1, k \rrbracket, \Phi_j \in \mathbb{Z}) = 0$ . Thus, the convergence in distribution directly follows from the previous lemma and the continuous mapping theorem.  $\square$

**Lemma 4.2.7.** *Let  $p \in [0, 1/2]$  and let  $B_n$  be a random variable following the binomial distribution of parameters  $n, p$ , for any  $n \in \mathbb{N}^*$ . Then  $(B_n)_{n \geq 1}$  satisfies the following properties:*

(i) *For all  $k \in \mathbb{N}^*$ ,*

$$\sup_{n \in \mathbb{N}^*} \mathbb{P}(B_n = k) \ll \frac{1}{\sqrt{k}}.$$

(ii) *Let  $1 \leq m \leq n$ . Let  $E$  an ensemble of integers between 1 and  $n$  such that:*

- *For all distinct  $j_1, j_2 \in E$ ,  $|j_1 - j_2| \geq m$ .*
- *For all  $j \in E$ ,  $j \geq m$ .*

*Then*

$$\mathbb{P}(B_n \in E) = \mathcal{O}\left(\frac{1}{\sqrt{m}}\right),$$

*where the  $\mathcal{O}(\frac{1}{\sqrt{m}})$  is independent of  $n$  and  $p$ .*

*Proof.* Proof of (i): Fix  $k \in \mathbb{N}^*$ . Using Stirling's formula and the fact that  $x^k(1-x)^{n-k}$

is maximal for  $x = \frac{k}{n}$ ,

$$\begin{aligned} \mathbb{P}(B_n = k) &= \binom{n}{k} p^k (1-p)^{n-k} \ll \frac{\left(\frac{n}{e}\right)^{n+1/2}}{\left(\frac{n-k}{e}\right)^{n-k+1/2} k!} \left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k} \\ &= \frac{1}{k!} \left(\frac{k}{e}\right)^k \sqrt{\frac{n}{n-k}} \\ &\ll \sqrt{\frac{n}{k(n-k)}}. \end{aligned}$$

Hence, if  $1 \leq k \leq \frac{3}{4}n$  we have  $\mathbb{P}(B_n = k) \ll \frac{1}{\sqrt{k}}$ . If  $\frac{3}{4}n \leq k \leq n$ , as  $x \rightarrow x^k(1-x)^{n-k}$  is increasing on  $[0, \frac{k}{n}]$  and  $p \leq \frac{1}{2} \leq \frac{k}{n}$  by hypothesis,

$$\mathbb{P}(B_n = k) \leq \binom{n}{\frac{n}{4}} \left(\frac{1}{2}\right)^n \ll \left(\frac{\frac{1}{2}}{\left(\frac{1}{4}\right)^{1/4} \left(\frac{3}{4}\right)^{3/4}}\right)^n = \left(\frac{2}{3^{3/4}}\right)^n \leq \left(\frac{2}{3^{3/4}}\right)^k \ll \frac{1}{\sqrt{k}}.$$

Proof of (ii): Fix  $n$  and denote  $f_k := \binom{n}{k} p^k (1-p)^{n-k}$ . For all  $0 \leq k < n$ , it is easy to check that

$$\frac{f_{k+1}}{f_k} < 1 \iff k > (n+1)p - 1,$$

so the sequence  $(f_k)_{k \in [0, n]}$  is increasing for  $k \leq k_0$ , and decreasing for  $k \geq k_0 + 1$ , where  $k_0 := \lfloor (n+1)p - 1 \rfloor$ .

If  $m \geq k_0 + 1$ , then

$$\mathbb{P}(B_n \in E) = \sum_{k \in E} f_k \leq \sum_{r=0}^{\lfloor \frac{n}{m} \rfloor - 1} f_{m+mr} \leq f_m + \frac{1}{m} \sum_{j=m}^{m \lfloor \frac{n}{m} \rfloor - 1} f_j \leq f_m + \frac{1}{m} \sum_{j=0}^n f_j = f_m + \frac{1}{m},$$

and by (i) we deduce  $\mathbb{P}(B_n \in E) \ll \frac{1}{\sqrt{m}}$ .

If  $m \leq k_0$ , then we look separately at the increasing part and the decreasing part: on the one hand,

$$\sum_{\substack{k \in E \\ k \leq k_0}} f_k \leq \sum_{r=0}^{\lfloor \frac{k_0}{m} \rfloor} f_{k_0 - mr} \leq f_{k_0} + \frac{1}{m} \sum_{j=k_0 - m \lfloor \frac{k_0}{m} \rfloor + 1}^{k_0} f_j \leq f_{k_0} + \frac{1}{m} \sum_{j=0}^{k_0} f_j,$$

and on the other hand,

$$\sum_{\substack{k \in E \\ k \geq k_0 + 1}} f_k \leq \sum_{r=0}^{\lfloor \frac{n - (k_0 + 1)}{m} \rfloor} f_{k_0 + 1 + mr} \leq f_{k_0 + 1} + \frac{1}{m} \sum_{j=k_0 + 1}^{k_0 + m \lfloor \frac{n - (k_0 + 1)}{m} \rfloor} f_j \leq f_{k_0 + 1} + \frac{1}{m} \sum_{j=k_0 + 1}^n f_j.$$

Thus,

$$\mathbb{P}(B_n \in E) = \sum_{k \in E} f_k \leq f_{k_0} + f_{k_0 + 1} + \frac{1}{m} \sum_{j=0}^n f_j = f_{k_0} + f_{k_0 + 1} + \frac{1}{m},$$

and by (i) we deduce  $\mathbb{P}(B_n \in E) \ll \frac{1}{\sqrt{k_0}} + \frac{1}{m} \ll \frac{1}{\sqrt{m}}$ .  $\square$

**Lemma 4.2.8.** *Assume  $\alpha$  is an irrational number of finite type  $\eta \geq 1$ . Let  $(a_j)_{j \geq 1}$  be any arbitrary sequence of positive real numbers. For all  $j$ , let  $E_j := \{\ell \in \mathbb{N}^* : \|\alpha\ell\| \leq a_j\}$ . Then for all  $\nu > \eta$ , there exists a number  $c_\nu$  depending only on  $\nu$  and  $a_1$ , such that for all  $j$ ,*

$$\sup_{n \in \mathbb{N}^*} \mathbb{P}_y(\ell_{n,j} \in E_j) \leq c_\nu a_j^{\frac{1}{2\nu}}.$$

*Proof.* Let  $\nu > \eta$ . From (4.16),

$$\exists C > 0, \forall \ell \in \mathbb{N}^*, \|\alpha\ell\| \geq \frac{C}{\ell^\nu}.$$

Let  $(a_j)_{j \geq 1}$  be a sequence of positive real numbers. For all  $j \in \mathbb{N}^*$ , it is easy to check that the set  $E_j := \{\ell \in \mathbb{N}^* : \|\alpha\ell\| \leq a_j\}$  satisfies

- $\forall \ell \in E_j, \ell \geq C^{1/\nu} a_j^{-1/\nu}$ .
- $\forall \ell_1 \neq \ell_2 \in E_j, |\ell_1 - \ell_2| \geq \left(\frac{C}{2}\right)^{1/\nu} a_j^{-1/\nu}$ .

Moreover, by construction, conditionally on  $y$  each random variable  $\ell_{n,j}$  follows a binomial distribution of parameters  $n, y_j$ . Since almost surely the  $y_j$  sum to one and  $y_1$  is the largest, then  $y_j \leq \frac{1}{2}$  for all  $j \neq 1$ , hence the previous lemma applies with  $m = \left(\frac{C}{2}\right)^{1/\nu} a_j^{-1/\nu}$  and gives for all  $j \geq 2$ ,

$$\mathbb{P}_y(\ell_{n,j} \in E_j) = \mathcal{O}\left(a_j^{\frac{1}{2\nu}}\right),$$

where the  $\mathcal{O}\left(a_j^{\frac{1}{2\nu}}\right)$  is independent of  $n$  and of the  $y_j, j \geq 2$ . In other words, there exists a number  $b_\nu$  depending only on  $\nu$  and  $y$ , such that for all  $j \geq 2$ ,

$$\sup_{n \in \mathbb{N}^*} \mathbb{P}_y(\ell_{n,j} \in E_j) \leq b_\nu a_j^{\frac{1}{2\nu}}.$$

Finally we get the result taking  $c_\nu := \max\left(b_\nu, a_1^{-\frac{1}{2\nu}}\right)$ . □

**Proposition 4.2.9.** *Conditionally on  $y$ , for all  $\varepsilon > 0$  and for all compact subsets  $K$  of  $\mathbb{C}$ ,*

$$\sup_{n \in \mathbb{N}^*} \mathbb{P}_y \left( \sup_{z \in K} \left| \prod_{\substack{j > k \\ \ell_{n,j} > 0}} \left( 1 + \frac{e^{2i\pi\alpha\ell_{n,j}}}{e^{2i\pi\alpha\ell_{n,j}} - 1} \left( e^{2i\pi z y_j^{(n)}} - 1 \right) \right) - 1 \right| \geq \varepsilon \right) \xrightarrow{k \rightarrow \infty} 0.$$

*Proof.* Let  $\varepsilon > 0$  and let  $K$  be a compact subset of  $\mathbb{C}$ . It suffices to show that conditionally on  $y$ ,

$$\sup_{n \in \mathbb{N}^*} \mathbb{P}_y \left( \sum_{\substack{j > k \\ \ell_{n,j} > 0}} \frac{y_j^{(n)}}{\|\alpha\ell_{n,j}\|} \geq \varepsilon \right) \xrightarrow{k \rightarrow \infty} 0,$$

Indeed, for all  $n, j$  such that  $\ell_{n,j} > 0$ ,

$$\left| \frac{e^{2i\pi\alpha\ell_{n,j}}}{e^{2i\pi\alpha\ell_{n,j}} - 1} \left( e^{2i\pi zy_j^{(n)}} - 1 \right) \right| \leq \frac{1}{\min(\{\alpha\ell_{n,j}\}, 1 - \{\alpha\ell_{n,j}\})} C_K y_j^{(n)} = \frac{C_K y_j^{(n)}}{\|\alpha\ell_{n,j}\|}$$

where  $C_K$  is a constant number that only depends on  $K$ .

Let  $n \geq 1$ . Let  $s, \rho \in (0, 1)$  such that  $r < \rho < s < 1$ . We have, denoting  $A :=$

$$\left\{ \sum_{\substack{j>k \\ \ell_{n,j}>0}} \frac{y_j^{(n)}}{\|\alpha\ell_{n,j}\|} \geq \varepsilon \right\} \text{ and } B_j := \{\|\alpha\ell_{n,j}\| \leq s^j, \ell_{n,j} > 0\},$$

$$\begin{aligned} \mathbb{P}_y(A) &= \mathbb{P}_y(A \cap B_{k+1}) + \mathbb{P}_y(A \cap B_{k+2}) + \dots + \mathbb{P}_y(A \cap (B_{k+1} \cup B_{k+2} \cup \dots)^c) \\ &\leq \left( \sum_{j>k} \mathbb{P}_y(B_j) \right) + \mathbb{P}_y(A \cap (B_{k+1} \cup B_{k+2} \cup \dots)^c). \end{aligned}$$

On the one hand, for all  $j$ ,  $\mathbb{P}_y(B_j) = \mathbb{P}_y(\ell_{n,j} \in E_j)$  with  $E_j := \{\ell \in \mathbb{N}^* : \|\alpha\ell\| \leq s^j\}$ , then it follows from Lemma 4.2.8 that  $\mathbb{P}_y(B_j) = \mathcal{O}\left(s^{\frac{j}{2\nu}}\right)$  independent of  $n$ , hence

$$\sup_{n \in \mathbb{N}^*} \sum_{j>k} \mathbb{P}_y(B_j) \xrightarrow[k \rightarrow \infty]{} 0.$$

On the other hand,

$$\begin{aligned} \mathbb{P}_y(A \cap (B_{k+1} \cup B_{k+2} \cup \dots)^c) &= \mathbb{P}_y(A \cap (\forall j > k, \|\alpha\ell_{n,j}\| > s^j \text{ or } \ell_{n,j} = 0)) \\ &\leq \mathbb{P}_y(\tilde{A}), \end{aligned}$$

with  $\tilde{A} := \left\{ \sum_{\substack{j>k \\ \ell_{n,j}>0}} \frac{y_j^{(n)}}{s^j} \geq \varepsilon \right\}$ , and furthermore denoting  $\tilde{B}_j := \{y_j^{(n)} \geq \rho^j\}$  it comes

$$\begin{aligned} \mathbb{P}_y(\tilde{A}) &\leq \left( \sum_{j>k} \mathbb{P}_y(\tilde{B}_j) \right) + \mathbb{P}_y(\tilde{A} \cap (\forall j > k, y_j^{(n)} < \rho^j)) \\ &\leq \left( \sum_{j>k} \frac{\mathbb{E}(y_j^{(n)} | y)}{\rho^j} \right) + \mathbb{P}_y\left(\sum_{j>k} \frac{\rho^j}{s^j} \geq \varepsilon\right) \\ &= \left( \sum_{j>k} \frac{y_j}{\rho^j} \right) + \mathbf{1}_{\sum_{j>k} \left(\frac{\rho}{s}\right)^j \geq \varepsilon} \end{aligned}$$

independent of  $n$ , hence

$$\sup_{n \in \mathbb{N}^*} \mathbb{P}_y(A \cap (B_{k+1} \cup B_{k+2} \cup \dots)^c) \xrightarrow[k \rightarrow \infty]{} 0.$$

Consequently,

$$\sup_{n \in \mathbb{N}^*} \mathbb{P}_y \left( \sum_{\substack{j > k \\ \ell_{n,j} > 0}} \frac{y_j^{(n)}}{\|\alpha \ell_{n,j}\|} \geq \varepsilon \right) \xrightarrow{k \rightarrow \infty} 0.$$

□

*Proof of Theorem 4.1.12 (ii).* Combining Proposition 4.2.6 and Proposition 4.2.9, we deduce that conditionally on  $y$ ,

$$\xi_{n,\alpha} \xrightarrow[n \rightarrow \infty]{} \tilde{\xi}_\infty.$$

In other words, for all functional  $F$  from  $\mathcal{C}(\mathbb{C}, \mathbb{C})$  to  $\mathbb{C}$ , continuous with respect to the topology of the uniform convergence on compact sets,

$$\mathbb{E}[F(\xi_{n,\alpha}) \mid y] \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \mathbb{E}[F(\tilde{\xi}_\infty) \mid y].$$

Hence by dominated convergence

$$\mathbb{E}[F(\xi_{n,\alpha})] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[F(\tilde{\xi}_\infty)].$$

□

### 4.3 Properties of the limiting function

In the following proposition we show that  $\tilde{\xi}_\infty$  has order one, in the sense of entire functions. The bound  $\mathcal{O}(\exp(c|z|))$  for this model can be compared to the one for the unitary case  $\mathcal{O}(\exp(c|z| \log |z|))$  established by Chhaibi, Najnudel and Nikeghbali in [30].

**Proposition 4.3.1.** *For all  $\varepsilon > 0$ , there exists a random number  $C_\varepsilon > 0$  such that for all  $z \in \mathbb{C}$ ,*

$$|\tilde{\xi}_\infty(z)| \leq C_\varepsilon e^{(2\pi+\varepsilon)|z|}. \quad (4.23)$$

*Proof.* The bound is clear for  $|z| < 1$  since  $\tilde{\xi}_\infty$  is locally bounded.

Assume  $|z| \geq 1$ . Let  $\eta > 0$ . Set  $k = \max\{j \geq 1 : 2\pi|z|y_j \geq \eta\} \vee 0$  and  $k_1 = \max\{j \geq 1 : 2\pi y_j \geq \eta\} \vee 0$ . Note that  $k \geq k_1$ , with  $k_1$  independent of  $z$ . We distinguish between two regimes depending on whether  $j$  is lower or greater than  $k$ .



- $j > k$ :

$$\begin{aligned}
 \prod_{j=k+1}^{+\infty} \left| \frac{e^{2i\pi zy_j} - u_j}{1 - u_j} \right| &\leq \prod_{j=k+1}^{+\infty} \left( 1 + \frac{2\pi|z|y_j}{|1 - u_j|} \exp(2\pi|z|y_j) \right) \\
 &\leq \prod_{j=k+1}^{+\infty} \left( 1 + \frac{2\pi|z|y_j}{|1 - u_j|} \exp(\eta) \right) \\
 &\leq \exp \left( 2\pi|z| \exp(\eta) \sum_{j=k+1}^{+\infty} \frac{y_j}{|1 - u_j|} \right) \\
 &\leq \exp \left( 2\pi|z| \exp(\eta) \sum_{j=k_1+1}^{+\infty} \frac{y_j}{|1 - u_j|} \right).
 \end{aligned}$$

Moreover we have seen from (4.20) that  $\sum_{j=1}^{+\infty} \frac{y_j}{|1 - u_j|} < \infty$ , and furthermore if  $\eta$  tends to zero then  $k_1$  goes to infinity, so we can chose  $\eta$  sufficiently close to 0 such that  $2\pi \exp(\eta) \sum_{j=k_1+1}^{+\infty} \frac{y_j}{|1 - u_j|} \leq \frac{\varepsilon}{2}$  and then

$$\prod_{j=k+1}^{+\infty} \left| \frac{e^{2i\pi zy_j} - u_j}{1 - u_j} \right| \leq \exp \left( \frac{\varepsilon}{2} |z| \right). \quad (4.24)$$

- $j \leq k$  (case to be considered only when  $k \neq 0$ ): As for all  $m \in \mathbb{N}^*$ ,

$$\left| \frac{e^{2i\pi zy_m} - u_m}{1 - u_m} \right| \leq \frac{e^{2\pi|z|y_m} + 1}{|1 - u_m|} \leq \frac{2e^{2\pi|z|y_m}}{|1 - u_m|},$$

then

$$\begin{aligned}
 \prod_{j=1}^k \left| \frac{e^{2i\pi zy_j} - u_j}{1 - u_j} \right| &\leq \left( \frac{2}{\min_{1 \leq j \leq k} |1 - u_j|} \right)^k \exp \left( 2\pi|z| \sum_{j=1}^k y_j \right) \\
 &\leq \left( \frac{2}{\min_{1 \leq j \leq k} |1 - u_j|} \right)^k e^{2\pi|z|}.
 \end{aligned} \quad (4.25)$$

It just remains to show that

$$\left( \frac{2}{\min_{1 \leq j \leq k} |1 - u_j|} \right)^k \leq C_\varepsilon \exp \left( \frac{\varepsilon}{2} |z| \right) \quad (4.26)$$

where  $C_\varepsilon$  is a random number depending on  $\varepsilon$  but not on  $z$ . Lemma 4.2.1 with

$\alpha = 3$  gives

$$\left( \frac{2}{\min_{1 \leq j \leq k} |1 - u_j|} \right)^k \leq \left( \frac{2}{C_1} k^3 \right)^k.$$

Moreover, using the assumption on the sequence  $(y_j)$ , there exists  $r \in (0, 1)$  and a (random) number  $C > 0$  such that for all  $j$ ,  $y_j \leq Cr^j$ , thus

$$k \leq \max \left\{ j \geq 1 : Cr^j \geq \frac{\eta}{2\pi|z|} \right\} \leq \frac{\log \left( \frac{2\pi C|z|}{\eta} \right)}{-\log r} \ll \log(|z| + 1)$$

Hence

$$\log \left( \left( \frac{2}{C_1} k^3 \right)^k \right) \ll \log(|z| + 1)(1 + \log \log(|z| + 1)) \underset{|z| \rightarrow \infty}{=} o(|z|),$$

which gives the existence of  $C_\varepsilon$ .

Consequently, (4.24), (4.25) and (4.26) jointly give (4.23).

□

Up to  $\varepsilon$ , the bound we provide in Proposition 4.3.1 is sharp. Indeed, we have the following result:

**Proposition 4.3.2.** *For all  $\varepsilon \in (0, 2\pi)$ , there exists a random number  $c_\varepsilon > 0$  such that for all  $x \geq 0$ ,*

$$|\tilde{\xi}_\infty(-ix)| \geq c_\varepsilon e^{(2\pi - \varepsilon)x}. \quad (4.27)$$

*Proof.* Let  $x \geq 0$ . For all  $k \in \mathbb{N}^*$ , if  $x \geq \frac{1}{2\pi y_k}$  then

$$\left| \frac{e^{2\pi x y_k} - u_k}{1 - u_k} \right| \geq 1 \geq \frac{1}{2} |e^{2\pi x y_k} - u_k| \geq \frac{1}{2} (e^{2\pi x y_k} - 1) \geq \frac{1}{4} e^{2\pi x y_k},$$

and if  $0 \leq x \leq \frac{1}{2\pi y_k}$  then

$$\left| \frac{e^{2\pi x y_k} - u_k}{1 - u_k} \right| \geq 1 \geq \frac{1}{e} e^{2\pi x y_k} \geq \frac{1}{4} e^{2\pi x y_k}.$$

Let  $\varepsilon \in (0, 2\pi)$ . Set  $p = \min\{k \geq 1 : \sum_{j=1}^k y_j \geq 1 - \frac{\varepsilon}{2\pi}\}$ . Then

$$\begin{aligned} |\tilde{\xi}_\infty(-ix)| &= \prod_{j=1}^{\infty} \left| \frac{e^{2\pi x y_j} - u_j}{1 - u_j} \right| \\ &\geq \prod_{j=1}^p \left| \frac{e^{2\pi x y_j} - u_j}{1 - u_j} \right| \\ &\geq \left(\frac{1}{4}\right)^p \exp\left(2\pi x \sum_{j=1}^p y_j\right) \\ &\geq \left(\frac{1}{4}\right)^p e^{(2\pi - \varepsilon)x}. \end{aligned}$$

□

## 4.4 More general central measures

In this section we treat the case of central measures such that  $\mathbb{P}\left(\sum_{j \geq 1} y_j < 1\right) > 0$ .

Let  $\lambda = (\lambda_j)_{j \geq 1}$  be a sequence of decreasing real numbers summing to  $\lambda_0 \in (0, 1)$ , and let  $E_\lambda = \bigsqcup_{j=1}^{\infty} \mathcal{C}_j \sqcup S$  be the disjoint union of circles  $\mathcal{C}_j$ , where for all  $j$ ,  $\mathcal{C}_j$  has perimeter  $\lambda_j$ , and of a segment  $S$  of length  $1 - \lambda_0$ . Let  $x = (x_k)_{k \geq 1} \in (\mathbb{E}_\lambda)^\infty$ .

**Proposition 4.4.1.** *One can define a virtual permutation  $\sigma_\infty(\lambda, x) = (\sigma_n(\lambda, x))_{n \geq 1}$ , where for all  $n$ , for all  $k \in \{1, \dots, n\}$ ,*

- *if  $x_k$  is on a circle, then the image of  $k$  by  $\sigma_n(\lambda, x)$  is the index of the first encountered point in  $\{x_1, \dots, x_n\}$  after  $x_k$  when exploring its circle counterclockwise (it is  $k$  itself if there is no other index  $j \leq n$  such that  $x_j$  is on the same circle as  $x_k$ ),*
- *if  $x_k$  is in  $S$ , then  $k$  is a fixed point of  $\sigma_n(\lambda, x)$ .*

Moreover, if  $\lambda$  follows any arbitrary distribution  $p$  on  $\nabla$ , and if conditionally on  $\lambda$  the points  $x_k$  are i.i.d following the uniform distribution on  $E_\lambda$ , then  $\sigma_\infty(\lambda, x)$  follows the central measure on  $\mathfrak{S}$  corresponding to  $p$ .

Let  $p$  be any probability measure on  $\nabla$ , and let  $(y_j)_{j \geq 1}$  be a random vector following the distribution  $p$ . Introduce the random variable  $y_0 := \sum_{j=1}^{+\infty} y_j$ .

For all  $n$  and  $j$ , denote  $\ell_{n,j} := \#\{k \in \{1, \dots, n\} : x_k \in \mathcal{C}_j\}$ , and  $p_n := \#\{k \in \{1, \dots, n\} : x_k \in S\}$ .

Let  $\alpha$  be an irrational number between 0 and 1. The expressions of  $\tilde{\xi}_n$  and  $\xi_{n,\alpha}$

defined at the beginning of the chapter become

$$\tilde{\xi}_n(z) = \left( \prod_{\substack{j \geq 1 \\ \ell_{n,j} > 0}} \frac{e^{2i\pi z y_j^{(n)}} - u_j}{1 - u_j} \right) \left( \prod_{k=1}^{p_n} \frac{e^{2i\pi \frac{z}{n}} - v_k}{1 - v_k} \right) \quad (4.28)$$

where the  $u_j$  and the  $v_k$  are independent random variables uniformly distributed on the unit circle, and

$$\xi_{n,\alpha} = \left( \prod_{\substack{j \geq 1 \\ \ell_{n,j} > 0}} \frac{e^{2i\pi(\frac{z}{n} + \alpha)\ell_{n,j}} - 1}{e^{2i\pi\alpha\ell_{n,j}} - 1} \right) \left( \frac{e^{2i\pi(\frac{z}{n} + \alpha)} - 1}{e^{2i\pi\alpha} - 1} \right)^{p_n}. \quad (4.29)$$

**Theorem 4.4.2.** *Assume that  $\sigma$  is generated by the coupling described above for a distribution with exponential decay  $p$  on  $\nabla$ . Then we have the following convergences in distribution:*

(i)

$$\tilde{\xi}_n(z) \xrightarrow{n \rightarrow \infty} \left( \prod_{j=1}^{+\infty} \frac{e^{2i\pi z y_j} - u_j}{1 - u_j} \right) e^{i\pi z(1-y_0)} \prod_{k \in \mathbb{Z}} \left( 1 - \frac{z}{w_k} \right)$$

where  $\{w_k : k \in \mathbb{Z}\}$  are points of a Poisson process with intensity  $1 - y_0$  on  $\mathbb{R}$  (if  $y_0 = 1$ , we make the convention  $\prod_{k \in \mathbb{Z}} \left( 1 - \frac{z}{w_k} \right) = 1$ ).

(ii) For all irrational number  $\alpha$  of finite type,

$$\xi_{n,\alpha}(z) \xrightarrow{n \rightarrow \infty} \left( \prod_{j=1}^{+\infty} \frac{e^{2i\pi z y_j} - u_j}{1 - u_j} \right) e^{i\pi z(1-y_0)\left(1 - \frac{i}{\tan(\pi\alpha)}\right)}.$$

*Remark 4.4.3.* In Theorem 4.4.2, the product  $\prod_{k \in \mathbb{Z}} \left( 1 - \frac{z}{w_k} \right)$  is not absolutely convergent.

It has to be understood as

$$\left( 1 - \frac{z}{w_0} \right) \prod_{k \geq 1} \left( 1 - \frac{z}{w_k} \right) \left( 1 - \frac{z}{w_{-k}} \right),$$

where the points of the Poisson process  $\{w_k : k \in \mathbb{Z}\}$  are labelled as follows:

$$\dots < w_{-2} < w_{-1} < 0 \leq w_0 < w_1 < w_2 < \dots$$

#### 4.4.1 Proof of Theorem 4.4.2 (i)

For all  $k \geq 1$ , write  $v_k = e^{2i\pi\Phi_k}$  where the  $\Phi_k$  are independent and uniformly distributed on  $[0, 1)$ . Then

$$\begin{aligned} \prod_{k=1}^{p_n} \frac{e^{2i\pi \frac{z}{n}} - v_k}{1 - v_k} &= \prod_{k=1}^{p_n} \frac{e^{2i\pi \frac{z}{n}} - e^{2i\pi\Phi_k}}{1 - e^{2i\pi\Phi_k}} \\ &= e^{i\pi z \frac{p_n}{n}} \prod_{k=1}^{p_n} \frac{\sin\left(\pi\left(\Phi_k - \frac{z}{n}\right)\right)}{\sin(\pi\Phi_k)} \\ &= e^{i\pi z \frac{p_n}{n}} \prod_{k=1}^{p_n} \frac{\Phi_k - \frac{z}{n}}{\Phi_k} \lim_{B \rightarrow +\infty} \prod_{0 < |j| \leq B} \frac{1 - \frac{\Phi_k - \frac{z}{n}}{j}}{1 - \frac{\Phi_k}{j}} \\ &= e^{i\pi z \frac{p_n}{n}} \prod_{k=1}^{p_n} \prod_{j \in \mathbb{Z}} \left(1 - \frac{z}{n(\Phi_k - j)}\right). \end{aligned}$$

*Remark 4.4.4.* The product in the last expression is not absolutely convergent. We write it like this for convenience of notation, and has to be understood as given by the expression above.

Denote by  $w_{n,k}$  the points  $n(\Phi_k - j)$  (defined for example as follows: for  $k$  from 1 to  $p_n$ ,  $w_{n,k} := n\Phi_k$ , and for all  $j \in \mathbb{Z}$ ,  $w_{n,k+jp_n} := w_{n,k} - jn$ ). The order of labelling for the points  $w_{n,k}$  does not matter for what we use in the sequel.

Let  $\mu_n := \sum_k \delta_{w_{n,k}}$  be the empirical measure associated with the point process of the  $w_{n,k}$ .

**Lemma 4.4.5.** *The empirical measure  $\mu_n$  converges vaguely to  $\mu_\infty$ , where  $\mu_\infty$  is the empirical measure associated with the points of a Poisson process with intensity  $1 - y_0$  on  $\mathbb{R}$ . The convergence holds for all compactly supported test functions from  $\mathbb{R}$  to  $\mathbb{C}$  (measurable but not necessarily continuous).*

*Proof.* Let  $f : \mathbb{R} \rightarrow \mathbb{C}$  such that  $\text{supp } f \subset [-M, M]$ ,  $M > 0$ . Let  $t \in \mathbb{R}$ . Let  $n > 2M$ . First note that by periodicity of the points  $w_{n,k}$ ,

$$\mathbb{E} \left( e^{it \sum_k f(w_{n,k})} \right) = \mathbb{E} \left( e^{it \sum_{k=1}^{X_n} f(\phi_k)} \right)$$

where  $X_n$  is a random variable which counts the number of the points  $w_{n,k}$  lying in  $[-M, M]$ , and where the  $\phi_k$  are i.i.d random variables uniformly chosen on  $[-M, M]$ ,

independently of  $X_n$ . Moreover,  $X_n$  is binomial of parameters  $p_n$  and  $2M/n$ . Hence,

$$\begin{aligned}
 \mathbb{E} \left( e^{it \sum_k f(w_{n,k})} \right) &= \mathbb{E} \left( \mathbb{E}(e^{itf(\phi_1)})^{X_n} \right) \\
 &= \sum_{k=0}^{p_n} \mathbb{E}(e^{itf(\phi_1)})^k \binom{p_n}{k} \left( \frac{2M}{n} \right)^k \left( 1 - \frac{2M}{n} \right)^{p_n-k} \\
 &= \left( 1 + \frac{2M}{n} (\mathbb{E}(e^{itf(\phi_1)}) - 1) \right)^{p_n} \\
 &= \left( 1 + \frac{1}{n} \int_{-M}^M (e^{itf(x)} - 1) dx \right)^{p_n} \\
 &\underset{n \rightarrow +\infty}{=} \exp \left( (1 - y_0) \int_{-M}^M (e^{itf(x)} - 1) dx \right) + o(1),
 \end{aligned}$$

since  $p_n/n \rightarrow 1 - y_0$  almost surely. Thus, the Fourier transform of  $\sum_k f(w_{n,k})$  converges to the Fourier transform of  $T := \sum_{x \in N} f(x)$ , where  $N$  is a homogeneous Poisson point process of parameter  $1 - y_0$  (for example the expression of the Fourier transform of  $T$  can be provided using the Campbell theorem), which gives the claim.  $\square$

**Proposition 4.4.6.** *For all  $A$ ,*

$$\prod_{|w_{n,k}| < A} \left( 1 - \frac{z}{w_{n,k}} \right) \underset{n \rightarrow +\infty}{\Rightarrow} \prod_{|w_k| < A} \left( 1 - \frac{z}{w_k} \right)$$

where  $\{w_k : k \in \mathbb{Z}\}$  are points of a Poisson process with intensity  $1 - y_0$  on  $\mathbb{R}$ .

*Proof.* Let  $A > 0$ . Let  $\mathcal{M}$  denote the space of locally finite measures of the form  $\sum_k \delta_{\alpha_k}$  for some arbitrary real numbers  $\alpha_k$ . Let  $F$  be the functional defined from  $\mathcal{M}$  to  $\mathcal{C}(\mathbb{C}, \mathbb{C})$  by

$$F \left( \sum_k \delta_{\alpha_k} \right) = \prod_k \left( 1 - \frac{z}{\alpha_k} \right) \mathbb{1}_{0 < |\alpha_k| < A}.$$

$F$  is continuous at every measure which does not charge  $-A$ ,  $0$ , and  $A$ . Since almost surely  $\mu_\infty$  (the empirical measure associated with the Poisson process  $\{w_k : k \in \mathbb{Z}\}$ ) does not charge these three points,  $F$  is continuous in  $\mu_\infty$ . By Lemma 4.4.5 and the continuous mapping theorem we deduce  $F(\mu_n) \rightarrow F(\mu_\infty)$ , which gives the claim.  $\square$

**Proposition 4.4.7.** *For all  $\varepsilon > 0$ , for all compact subsets  $K$  of  $\mathbb{C}$ ,*

$$\sup_{n \in \mathbb{N}^*} \mathbb{P} \left( \sup_{z \in K} \left| \prod_{|w_{n,k}| \geq A} \left( 1 - \frac{z}{w_{n,k}} \right) - 1 \right| \geq \varepsilon \right) \xrightarrow{A \rightarrow +\infty} 0.$$

*Proof.* Let  $K$  a compact subset of  $\mathbb{C}$  of diameter  $D$  for the uniform norm, and let  $A > 2D$ .

For all  $z \in K$ ,

$$\prod_{|w_{n,k}| \geq A} \left(1 - \frac{z}{w_{n,k}}\right) = \exp \left( -z \sum_{|w_{n,k}| \geq A} \frac{1}{w_{n,k}} \right) \exp \left( \mathcal{O}_K \left( \sum_{|w_{n,k}| \geq A} \frac{1}{w_{n,k}^2} \right) \right). \quad (4.30)$$

The sum  $\sum_{|w_{n,k}| \geq A} \frac{1}{w_{n,k}}$  is not absolutely convergent. Let  $B > A$ . Integrating by parts,

$$\begin{aligned} \sum_{A \leq |w_{n,k}| \leq B} \frac{1}{w_{n,k}} &= \int_{-B}^{-A} \frac{1}{x} d\mu_n(x) + \int_A^B \frac{1}{x} d\mu_n(x) \\ &= \frac{\mu_n[A, B] - \mu_n[-B, -A]}{B} + \int_A^B \frac{\mu_n[A, x] - \mu_n[-x, -A]}{x^2} dx. \end{aligned}$$

As for all  $a, b \in \mathbb{R}$ ,  $\left| \mu_n[a, b] - \frac{(b-a)}{n} p_n \right| \leq 2p_n$ , then  $\lim_{B \rightarrow +\infty} \frac{\mu_n[A, B] - \mu_n[-B, -A]}{B} = 0$ , and we get

$$\sum_{|w_{n,k}| \geq A} \frac{1}{w_{n,k}} = \int_A^{+\infty} \frac{\mu_n[A, x] - \mu_n[-x, -A]}{x^2} dx.$$

Hence

$$\mathbb{E} \left| \sum_{|w_{n,k}| \geq A} \frac{1}{w_{n,k}} \right| \leq \int_A^{+\infty} \frac{1}{x^2} \mathbb{E} |\mu_n[A, x] - \mu_n[-x, -A]| dx,$$

with for all  $x > A$ , by Cauchy-Schwarz inequality,

$$\begin{aligned} (\mathbb{E} |\mu_n[A, x] - \mu_n[-x, -A]|)^2 &\leq \mathbb{E} ((\mu_n[A, x] - \mu_n[-x, -A])^2) \\ &= \text{Var}(\mu_n[A, x] - \mu_n[-x, -A]) \\ &\leq 2(\text{Var}(\mu_n[A, x]) + \text{Var}(\mu_n[-x, -A])) \\ &= 4\text{Var}(\mu_n[A, x]) \end{aligned}$$

since  $\mu_n[A, x]$  and  $\mu_n[-x, -A]$  are equally distributed (consequence of the fact that the  $\Phi_k$  are uniformly distributed on  $[0, 1)$ ).

As each interval of the form  $[jn, (j+1)n)$  contains exactly  $p_n$  points, and these points are uniformly distributed, then  $\text{Var}(\mu_n[A, x])$  is the variance of a binomial random variable of parameters  $p_n$  and  $\left\{ \frac{x-A}{n} \right\}$ , that is

$$\begin{aligned} \text{Var}(\mu_n[A, x]) &= p_n \left\{ \frac{x-A}{n} \right\} \left( 1 - \left\{ \frac{x-A}{n} \right\} \right) \\ &\leq p_n \frac{x-A}{n} \\ &\leq x - A. \end{aligned}$$

We deduce

$$\mathbb{E} \left| \sum_{|w_{n,k}| \geq A} \frac{1}{w_{n,k}} \right| \leq 2 \int_A^{+\infty} \frac{\sqrt{x-A}}{x^2} dx = \frac{\pi}{\sqrt{A}} \xrightarrow{A \rightarrow +\infty} 0.$$

Moreover,

$$\mathbb{E} \sum_{|w_{n,k}| \geq A} \frac{1}{w_{n,k}^2} = \int_{-\infty}^{-A} \frac{1}{x^2} \frac{p_n}{n} dx + \int_A^{+\infty} \frac{1}{x^2} \frac{p_n}{n} dx \leq \frac{2}{A} \xrightarrow{A \rightarrow +\infty} 0.$$

From (4.30) we deduce that  $z \mapsto \prod_{|w_{n,k}| \geq A} \left(1 - \frac{z}{w_{n,k}}\right)$  converges in probability to  $z \mapsto 1$  on every compact sets as  $A$  goes to  $+\infty$ , uniformly in  $n$ , which gives the claim.  $\square$

Proposition 4.4.6 and Proposition 4.4.7 together show

$$\prod_{k=1}^{p_n} \prod_{j \in \mathbb{Z}} \left(1 - \frac{z}{n(\Phi_k - j)}\right) \xrightarrow{n \rightarrow +\infty} \prod_{k \in \mathbb{Z}} \left(1 - \frac{z}{w_k}\right)$$

(see remark of Theorem 4.4.2 for the sense given to this last non-absolutely convergent product), hence

$$\prod_{k=1}^{p_n} \frac{e^{2i\pi \frac{z}{n}} - v_k}{1 - v_k} \xrightarrow{n \rightarrow +\infty} e^{i\pi z(1-y_0)} \prod_{k \in \mathbb{Z}} \left(1 - \frac{z}{w_k}\right).$$

Finally, in the same manner as in the proof of point (i) of Theorem 4.1.12, we prove the almost sure convergence of  $\prod_{\substack{j \geq 1 \\ \ell_{n,j} > 0}} \frac{e^{2i\pi z y_j^{(n)}} - u_j}{1 - u_j}$  to  $\prod_{j=1}^{+\infty} \frac{e^{2i\pi z y_j} - u_j}{1 - u_j}$ . Conditionally on  $y$

and  $(u_j)_{j \geq 1}$ , Slutsky's theorem applies on the functional space  $\mathcal{C}(\mathbb{C}, \mathbb{C})$ , which allows to conclude using the dominated convergence theorem.

#### 4.4.2 Proof of Theorem 4.4.2 (ii)

The proof of point (ii) is much simpler. Indeed, it suffices to see that for all  $n$  and for all  $z$  in any compact subset  $K$  of  $\mathbb{C}$ ,

$$\begin{aligned} \left( \frac{e^{2i\pi(\frac{z}{n} + \alpha)} - 1}{e^{2i\pi\alpha} - 1} \right)^{p_n} &= \left( 1 + \frac{e^{2i\pi\alpha}}{e^{2i\pi\alpha} - 1} \left( e^{2i\pi \frac{z}{n}} - 1 \right) \right)^{p_n} \\ &= \exp \left( p_n \left[ \frac{2i\pi e^{2i\pi\alpha}}{e^{2i\pi\alpha} - 1} \frac{z}{n} + \mathcal{O} \left( \frac{1}{n^2} \right) \right] \right) \\ &= \exp \left( \frac{p_n}{n} \frac{2i\pi z e^{2i\pi\alpha}}{e^{2i\pi\alpha} - 1} + \mathcal{O} \left( \frac{1}{n} \right) \right) \\ &= \exp \left( (1 - y_0) \frac{2i\pi z e^{2i\pi\alpha}}{e^{2i\pi\alpha} - 1} \right) + o(1) \end{aligned}$$

uniformly in  $z \in K$ , and for all  $n$  large enough depending on  $\alpha$  and  $K$ .

Finally, simplifying  $\frac{2ie^{2i\pi\alpha}}{e^{2i\pi\alpha} - 1} = \frac{e^{i\pi\alpha}}{\sin(\pi\alpha)} = \frac{1}{\tan(\pi\alpha)} + i$  gives the claim.



### 4.4.3 Properties of the limiting functions

**Lemma 4.4.8.** *Let  $\varepsilon > 0$ . Almost surely, for all  $k \in \mathbb{Z}$ ,*

$$w_k = \frac{k}{1 - y_0} + \mathcal{O}\left(k^{\frac{1}{2} + \varepsilon}\right). \quad (4.31)$$

We omit the proof of this lemma since it is a classical result on Poisson processes.

**Proposition 4.4.9.** *For all  $\varepsilon > 0$ , there exists a random number  $C > 0$  such that for all  $z \in \mathbb{C}$ ,*

$$|\tilde{\xi}_\infty(z)| \leq e^{C|z| \log(2+|z|)}. \quad (4.32)$$

*Proof.* First, following the same reasoning as in Section 4.3, it is easy to check that for all  $\varepsilon > 0$ , there exists a random number  $C_\varepsilon > 0$  such that for all  $z \in \mathbb{C}$ ,

$$\left| \left( \prod_{j=1}^{+\infty} \frac{e^{2i\pi z y_j} - u_j}{1 - u_j} \right) \right| \leq C_\varepsilon e^{(2\pi y_0 + \varepsilon)|z|}.$$

Hence there exists a random number  $c > 0$  such that for all  $z \in \mathbb{C}$ ,

$$\left| \left( \prod_{j=1}^{+\infty} \frac{e^{2i\pi z y_j} - u_j}{1 - u_j} \right) e^{i\pi z(1-y_0)} \right| \leq e^{c|z|}.$$

Thus, it is enough to show that there exists a random number  $C > 0$  such that for all  $z \in \mathbb{C}$ ,

$$\left| \prod_{k \in \mathbb{Z}} \left( 1 - \frac{z}{w_k} \right) \right| \leq e^{C|z| \log(2+|z|)}. \quad (4.33)$$

To this end, we distinguish between two regimes of  $k \neq 0$  in this product:  $|k| \geq |z|$ , and  $1 \leq |k| \leq |z|$ . For the first regime, using the previous lemma with  $\varepsilon = \frac{1}{3}$  gives

$$\left( 1 - \frac{z}{w_k} \right) \left( 1 - \frac{z}{w_{-k}} \right) = 1 + \mathcal{O}\left( \frac{|z|k^{\frac{1}{2} + \frac{1}{3}} + |z|^2}{k^2} \right),$$

hence

$$\begin{aligned} \left| \prod_{k \geq |z|} \left( 1 - \frac{z}{w_k} \right) \left( 1 - \frac{z}{w_{-k}} \right) \right| &\leq \exp \left( \mathcal{O} \left( \sum_{k \geq |z|} |z|k^{-\frac{7}{6}} + |z|^2 k^{-2} \right) \right) \\ &= \exp(\mathcal{O}(|z|)). \end{aligned} \quad (4.34)$$

For the second regime, as  $\left| \frac{w_k}{k} \right|$  is almost surely bounded from below (since  $\frac{w_k}{k} = \frac{1}{1-y_0} + \mathcal{O}(k^{-\frac{1}{2} + \varepsilon})$ ), we have

$$1 - \frac{z}{w_k} = 1 + \mathcal{O}\left( \left| \frac{z}{k} \right| \right)$$

and it follows

$$\left| \prod_{1 \leq k < |z|} \left(1 - \frac{z}{w_k}\right) \left(1 - \frac{z}{w_{-k}}\right) \right| \leq \exp \left( \mathcal{O} \left( \sum_{1 \leq k < |z|} \frac{|z|}{k} \right) \right) \quad (4.35)$$

$$= \exp(\mathcal{O}(|z| \log(2 + |z|))).$$

Furthermore,

$$\left| 1 - \frac{z}{w_0} \right| \leq \exp \left( \frac{|z|}{|w_0|} \right) = \exp(\mathcal{O}(|z|)) \quad (4.36)$$

since  $w_0 \neq 0$  almost surely. Combining (4.34), (4.35) and (4.36), we deduce the existence of a random number  $C > 0$  such that for all  $z \in \mathbb{C}$  we have (4.33), and the proof is complete.  $\square$

**Proposition 4.4.10.** *For all  $\varepsilon > 0$ , there exists a random number  $C_\varepsilon > 0$  such that for all  $z \in \mathbb{C}$ ,*

$$|\xi_{\infty, \alpha}(z)| \leq C_\varepsilon e^{(\varepsilon + 2\pi(y_0 + (1-y_0)t_\alpha)|z|)}, \quad (4.37)$$

where  $t_\alpha = \frac{1}{2 \sin(\pi \alpha)} \in (\frac{1}{2}, +\infty)$ .

The proof of the last proposition is very similar to the one for the case  $\sum_{j=1}^{+\infty} y_j = 1$  almost surely. We omit it here and refer to Section 4.3.

## 5. Conclusion et perspectives

Le travail de cette thèse nous a permis de mieux comprendre la répartition des valeurs propres de matrices de permutation.

De nombreuses ressemblances avec l'ensemble circulaire unitaire (CUE) en ressortent. En effet, premièrement, concernant les nombres de valeurs propres dans des arcs macroscopiques ou mésoscopiques, le comportement de leur variance est aussi logarithmique, et les fluctuations sont asymptotiquement gaussiennes avec une structure de covariance similaire pour le cas de plusieurs arcs fixés à extrémités typiques. Ceci avait déjà été observé par Wieand pour la loi uniforme avec des arcs fixés, et notre étude dans le Chapitre 2 a permis de voir que l'analogie avec le CUE reste valable pour les lois d'Ewens et des arcs mésoscopiques.

Deuxièmement, à l'échelle microscopique, le travail présenté au Chapitre 3 montre particulièrement que le nombre de points du processus ponctuel limite sur un intervalle vérifie un théorème central limite lorsque la longueur de l'intervalle tend vers l'infini, et ceci pour de nombreux cas d'intervalles. Pour le CUE le processus ponctuel limite correspondant est un processus déterminantal de noyau sinus, et il a été montré par Costin et Lebowitz [31] que le nombre de points vérifie aussi un théorème central limite. Enfin, nos résultats du Chapitre 4 témoignent eux aussi de similarités avec le CUE, tant ils sont proches de certains résultats établis par Chhaibi, Najnudel et Nikeghbali [30]. Pour une grande famille de lois sur le groupe symétrique, nous avons en effet obtenu que le polynôme caractéristique normalisé converge vers une fonction holomorphe qui partage certaines mêmes propriétés que pour le CUE. De plus, nous avons montré que la notion d'isométries virtuelles introduite par Najnudel et Nikeghbali dans le cadre des matrices unitaires s'applique aussi à des matrices de permutation modifiées, ce qui a permis d'avoir une convergence presque sûre.

Il pourrait être intéressant de creuser davantage l'étude des liens entre le CUE et les matrices de permutation (avec ou sans modifications).

Un grand nombre d'autres directions sont envisageables en perspectives de nos travaux. Nous en donnons quelques unes en les mettant en relief avec des travaux récents.

### Lois plus générales

Plusieurs méthodes que nous avons employées dans cette thèse sont liées au couplage de Feller, qui est assez spécifique aux lois d'Ewens. Si l'on veut s'intéresser à des lois d'Ewens généralisées (c'est-à-dire lorsque le paramètre  $\theta$  est autorisé à changer selon les tailles de cycles), il semble difficile de trouver un équivalent du couplage de Feller permettant

d'étendre facilement les résultats. Par exemple pour ces lois d'Ewens généralisées on trouve dans la littérature des méthodes d'analyse complexe avec la considération de fonctions génératrices [73], [73], [82].

Si l'on veut toutefois essayer de trouver d'autres lois auxquelles les techniques relatives au couplage de Feller s'appliquent, cela est possible et Zhao [94] donne des conditions suffisantes que doivent satisfaire ces lois.

Cependant, le couplage de Feller n'est pas indispensable pour l'étude des permutations aléatoires, même dans le cadre des lois d'Ewens. Par exemple, Féray [41] s'appuie sur des méthodes de moments et de cumulants pour étudier certaines statistiques de permutations suivant des lois d'Ewens. Son étude fournit beaucoup d'outils et donc autant de possibilités d'exploration.

Un autre procédé que nous avons utilisé dans cette thèse est la représentation de la loi GEM comme produit de variables aléatoires Beta indépendantes. Une généralisation immédiate est connue sous le nom de *stick breaking process*, qui consiste à générer une suite de variables aléatoires  $x = (x_i)_{i \geq 0}$  de  $[0, 1]$  sommant à 1 à partir de variables i.i.d de loi arbitraire. Or comme nous l'avons vu dans le Chapitre 4, chaque réalisation de  $x$ , réordonnée de manière décroissante, correspond à une unique mesure centrale sur l'espace des permutations virtuelles. Le stick breaking process donne donc accès à une grande famille de lois de tirage des permutations, incluant les lois d'Ewens. Gnedin, Iksanov et Marynych [43] ont exploité cette méthode pour généraliser la loi d'Erdős et Turán. Citons aussi les travaux de Pitman et Tang [77], Duchamps, Pitman et Tang [37], ou encore ceux de Cristali, Ranjan, Steinberg, Beckman, Durrett, Junge et Nolen [32] utilisant cette méthode dans le cadre de processus de renouvellement/régénératifs liés à des permutations.

## Statistiques linéaires

Ben Arous et Dang [13] ont étudié les statistiques linéaires du spectre de matrices de permutation à échelle macroscopique, mais à notre connaissance il n'y a jamais eu d'étude semblable réalisée à d'autres échelles concernant les matrices de permutation. Qu'en est-il alors pour les échelles mésoscopique et microscopique ?

À l'échelle mésoscopique, Lambert obtient en particulier des résultats asymptotiques pour des statistiques linéaires régulières sur certains processus déterminantaux liés à des ensembles de matrices aléatoires unitairement invariants [62], ainsi que pour des régularisations d'indicatrices d'intervalles [63]. Par ailleurs, Berggren et Duits [16] étudient des statistiques linéaires régulières du *thinned circular unitary ensemble*, c'est-à-dire le CUE où chaque valeur propre est enlevée avec probabilité  $\gamma \in ]0, 1[$  de manière indépendante. Il serait intéressant de voir si leurs résultats ont des versions similaires pour les matrices de permutation.

À l'échelle microscopique, les méthodes que nous avons utilisées au Chapitre 3 semblent pouvoir s'appliquer à une grande famille de fonctions  $f$  à supports compacts. Plus spécifiquement, pour les matrices de permutation sans modification, la méthode du couplage avec un processus de Poisson invariant d'échelle repose principalement sur le fait que  $f$  est bornée, et que  $x \mapsto \min(f(x), 1)/x$  est intégrable sur  $\mathbb{R}$  afin d'appliquer le

théorème de Campbell. Il serait intéressant de voir si on peut dégager des comportements typiques des fluctuations asymptotiques de ces statistiques linéaires.

## Écarts entre valeurs propres

À la fin du Chapitre 2 nous commençons une étude de l'écart maximal et minimal entre deux valeurs propres distinctes consécutives, en obtenant des résultats de tensions de variables aléatoires. Il serait intéressant de voir si on peut affiner cette étude avec des convergences en lois lorsque la taille  $n$  des matrices tend vers l'infini. Par exemple un résultat obtenu par Vinson pour le CUE est que le plus petit écart multiplié par  $n^{4/3}$  converge en loi vers une variable aléatoire dont la fonction de densité est  $x \mapsto e^{-x^3}$ , lorsque  $n$  tend vers l'infini. Ben Arous et Bourgade [12] poursuivent l'étude pour le CUE et le GUE et en exhibent des liens avec les écarts entre les zéros de la fonction zeta de Riemann.

Par ailleurs, dans la continuité du Chapitre 3, il pourrait être intéressant d'étudier l'écart maximal/minimal entre deux points distincts consécutifs dans un intervalle  $J$  du processus ponctuel limite, lorsque la longueur de  $J$  tend vers l'infini. Pour ce faire, un travail préliminaire consisterait peut-être à s'inspirer des travaux récents de Pitman [78] sur les écarts entre les coordonnées d'un vecteur aléatoire de loi GEM.

## Polynômes caractéristiques

Dans le Chapitre 4, nous n'avons pas poussé l'étude du polynôme caractéristique aussi loin que l'ont fait Chhaibi, Najnudel et Nikeghbali pour le CUE [30]. Par exemple, on pourrait étudier des ratios en différents points du polynôme caractéristique, ou s'intéresser aux propriétés de la dérivée logarithmique des fonctions limites. On pourrait aussi essayer d'obtenir des résultats pour le polynôme caractéristique à échelle mésoscopique.

Par ailleurs, les résultats que nous avons établis sur les polynômes caractéristiques sont valables pour une grande famille de mesures centrales : celles associées à des mesures  $p$  sur  $\nabla$  vérifiant une hypothèse de décroissance exponentielle. Une question naturelle est alors de se demander ce qu'il se passe lorsque  $p$  n'est pas à décroissance exponentielle. Par exemple si  $p$  est à décroissance polynomiale, que sait-on faire dans ce cas ?

Une toute autre question serait de savoir si pour les matrices de permutation modifiées (ou non modifiées) on a des résultats analogues à ceux de Chhaibi, Najnudel et Madaule [29] concernant le maximum du logarithme du polynôme caractéristique du C $\beta$ E en les points du cercle unité. En particulier, en notant  $N_n^I$  le nombre de valeurs propres dans  $I$  d'extrémités  $e^{2i\pi\alpha}$  et  $e^{2i\pi\beta}$  d'une matrice de permutation aléatoire (modifiée ou non) de taille  $n$ , existe-t-il des constantes  $\kappa_1, \kappa_2$  telles que la suite de variables aléatoires

$$\left( \sup_{\alpha, \beta} |N_n^I - \mathbb{E}(N_n^I)| - (\kappa_1 \ln n + \kappa_2 \ln \ln n) \right)_{n \geq 2}$$

est tendue ?

Terminons par dire quelques mots sur des directions que nous n'avons pas explorées pendant cette thèse mais ayant rapport avec des permutations aléatoires.

## Permutations avec contraintes

Récemment, beaucoup d'études probabilistes ont été menées sur des sous-ensembles de groupes symétriques.

Par exemple, Betz et Schäfer [17], et Betz et Zeindler [18] étudient des permutations aléatoires sans grands cycles.

Cela peut être aussi des contraintes au niveau de certains motifs. Par exemple, Janson [54] s'intéresse à des permutations évitant le motif 132. Parallèlement, Bassino, Bouvel, Féray, Guérin, Maazoun et Pierrot [11], [10], [64] développent des résultats faisant intervenir des mesures sur le carré unité appelées *permutons*. Il est connu que des permutons peuvent s'obtenir comme des limites en loi de mesures associées à des permutations aléatoires évitant certains motifs. Toutefois, à notre connaissance, le point de vue des suites de matrices de permutation n'a pas été utilisé dans ce contexte. Cela ne paraît en fait pas très surprenant, car les ensembles de permutations évitant certains motifs ne sont généralement pas stables par conjugaison, ce qui amène donc à la considération de mesures non centrales.

## Probabilités libres, produits et sommes de permutations aléatoires, théorie des graphes

Dans cette thèse nous n'avons pas eu recours à des notions de théorie des représentations de groupes ou de probabilités libres, bien que les aspects combinatoires soient très présents dans ces théories, et que les probabilités libres donnent souvent un cadre adapté à l'étude de certains problèmes en matrices aléatoires.

Biane [20] met en avant la théorie des probabilités libres pour l'étude des représentations sur le groupe symétrique. Par ailleurs, dans [19] il établit en particulier une approximation de la loi semi-circulaire par une somme de transpositions vues comme des opérateurs. Le modèle par permutations qu'il fournit est développé par Benaych-Georges et Nechita [15] qui font un lien avec la théorie des probabilités classiques. Citons aussi Neagu [70] et Male [65] qui utilisent des matrices de permutation en théorie des matrices aléatoires et probabilités libres.

Il nous aurait sans doute fallu explorer dans ces directions si nous avions eu affaire à des produits ou des sommes de permutations aléatoires.

Des produits de permutations peuvent avoir une interprétation en terme de mots, ou bien de marches aléatoires. Hultman [53] les utilise par exemple à travers la théorie de représentations pour étudier certaines statistiques sur le groupe symétrique. En s'inspirant en particulier des travaux de Nica [72], Neagu [71], et Yakimiv [91], Benaych-Georges [14] considère des mots libres fixés  $w$  où chaque lettre  $g_i$  dans  $w$  est remplacée par une permutation aléatoire  $s_i$  de  $[n]$  ayant ses tailles de cycles dans un certain ensemble  $A_i$  d'entiers fixé. Il montre notamment que dans de nombreux cas sur les  $w$  et les  $A_i$ , les

nombre de cycles de taille  $k$  de la permutation associée à  $w$  convergent conjointement en loi vers des Poisson indépendantes de paramètres  $1/k$ .

Enfin, des sommes de matrices de permutation peuvent avoir une interprétation en terme de graphes. Nous renvoyons par exemple vers les travaux récents et variés de Johnson et Pal [55], Bordenave et Collins [22], et Basak, Cook et Zeitouni [9].





# Bibliography

- [1] Mélisande Albert. *Tests d'indépendance par bootstrap et permutation: étude asymptotique et non-asymptotique. Application en neurosciences*. PhD thesis, Nice, 2015.
- [2] Richard Arratia. On the central role of scale invariant Poisson processes on  $(0, \infty)$ . In *Microsurveys in discrete probability (Princeton, NJ, 1997)*, volume 41 of *DI-MACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 21–41. Amer. Math. Soc., Providence, RI, 1998.
- [3] Richard Arratia, A. D. Barbour, and Simon Tavaré. Poisson process approximations for the Ewens sampling formula. *Ann. Appl. Probab.*, 2(3):519–535, 1992.
- [4] Richard Arratia, A. D. Barbour, and Simon Tavaré. *Logarithmic combinatorial structures: a probabilistic approach*. EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich, 2003.
- [5] Richard Arratia, A. D. Barbour, and Simon Tavaré. A tale of three couplings: Poisson-Dirichlet and GEM approximations for random permutations. *Combin. Probab. Comput.*, 15(1-2):31–62, 2006.
- [6] Richard Arratia and Simon Tavaré. Limit theorems for combinatorial structures via discrete process approximations. *Random Structures Algorithms*, 3(3):321–345, 1992.
- [7] Valentin Bahier. On the number of eigenvalues of modified permutation matrices in mesoscopic intervals. *Journal of Theoretical Probability*, Dec 2017.
- [8] Jinho Baik, Percy Deift, and Kurt Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, 12(4):1119–1178, 1999.
- [9] Anirban Basak, Nicholas Cook, and Ofer Zeitouni. Circular law for the sum of random permutation matrices. *arXiv preprint arXiv:1705.09053*, 2017.
- [10] Frédérique Bassino, Mathilde Bouvel, Valentin Féray, Lucas Gerin, Mickaël Maa-zoun, and Adeline Pierrot. Universal limits of sunstitution-closed permutation classes. *arXiv preprint arXiv:1706.08333*, 2017.
- [11] Frédérique Bassino, Mathilde Bouvel, Valentin Féray, Lucas Gerin, and Adeline Pierrot. The brownian limit of separable permutations. *arXiv preprint arXiv:1602.04960*, 2016.

- [12] Gérard Ben Arous and Paul Bourgade. Extreme gaps between eigenvalues of random matrices. *Ann. Probab.*, 41(4):2648–2681, 2013.
- [13] Gérard Ben Arous and Kim Dang. On fluctuations of eigenvalues of random permutation matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(2):620–647, 2015.
- [14] Florent Benaych-Georges. Cycles of free words in several independent random permutations with restricted cycle lengths. *Indiana Univ. Math. J.*, 59(5):1547–1586, 2010.
- [15] Florent Benaych-Georges and Ion Nechita. A permutation model for free random variables and its classical analogue. *Pacific J. Math.*, 242(1):33–51, 2009.
- [16] Tomas Berggren and Maurice Duits. Mesoscopic fluctuations for the thinned circular unitary ensemble. *Math. Phys. Anal. Geom.*, 20(3):Art. 19, 40, 2017.
- [17] Volker Betz and Helge Schäfer. The number of cycles in random permutations without long cycles is asymptotically Gaussian. *ALEA Lat. Am. J. Probab. Math. Stat.*, 14(1):427–444, 2017.
- [18] Volker Betz, Helge Schäfer, and Dirk Zeindler. Random permutations without macroscopic cycles. *arXiv preprint arXiv:1712.04738*, 2017.
- [19] Philippe Biane. Permutation model for semi-circular systems and quantum random walks. *Pacific J. Math.*, 171(2):373–387, 1995.
- [20] Philippe Biane. Representations of symmetric groups and free probability. *Adv. Math.*, 138(1):126–181, 1998.
- [21] Nathaniel Blair-Stahn. Random permutation matrices. 2000.
- [22] Charles Bordenave and Benoît Collins. Eigenvalues of random lifts and polynomial of random permutations matrices. *arXiv preprint arXiv:1801.00876*, 2018.
- [23] Paul Bourgade. Mesoscopic fluctuations of the zeta zeros. *Probab. Theory Related Fields*, 148(3-4):479–500, 2010.
- [24] Paul Bourgade, Chris Hughes, Ashkan Nikeghbali, and Marc Yor. The characteristic polynomial of a random unitary matrix: a probabilistic approach. *Duke Math. J.*, 145(1):45–69, 2008.
- [25] Paul Bourgade, Joseph Najnudel, and Ashkan Nikeghbali. A unitary extension of virtual permutations. *Int. Math. Res. Not. IMRN*, (18):4101–4134, 2013.
- [26] Paul Bourgade, Ashkan Nikeghbali, and Alain Rouault. Ewens measures on compact groups and hypergeometric kernels. In *Séminaire de Probabilités XLIII*, volume 2006 of *Lecture Notes in Math.*, pages 351–377. Springer, Berlin, 2011.
- [27] Djalil Chafaï, Yan Doumerc, and Florent Malrieu. Processus des restaurants chinois et loi d’Ewens. *Revue de Mathématiques Spéciales (RMS)*, 123(3):56–74, 2013.

- [28] Reda Chhaibi, Emma Hovhannisyan, Joseph Najnudel, Ashkan Nikeghbali, and Brad Rodgers. A limiting characteristic polynomial of some random matrix ensembles. *arXiv preprint arXiv:1707.09956*, 2017.
- [29] Reda Chhaibi, Thomas Madaule, and Joseph Najnudel. On the maximum of the  $C\beta E$  field. *arXiv preprint arXiv:1607.00243*, 2016.
- [30] Reda Chhaibi, Joseph Najnudel, and Ashkan Nikeghbali. The circular unitary ensemble and the Riemann zeta function: the microscopic landscape and a new approach to ratios. *Invent. Math.*, 207(1):23–113, 2017.
- [31] Ovidiu Costin and Joel L. Lebowitz. Gaussian fluctuation in random matrices. *Phys. Rev. Lett.*, 75(1):69–72, 1995.
- [32] Irina Cristali, Vinit Ranjan, Jake Steinberg, Erin Beckman, Rick Durrett, Matthew Junge, and James Nolen. Block sizes in two families of random permutations. *arXiv preprint arXiv:1708.05626*, 2017.
- [33] Kim Dang. *The spectrum of random permutation matrices*. PhD thesis, University of Zurich, 2012.
- [34] Kim Dang and Dirk Zeindler. The characteristic polynomial of a random permutation matrix at different points. *Stochastic Process. Appl.*, 124(1):411–439, 2014.
- [35] Paul-Olivier Dehaye and Dirk Zeindler. On averages of randomized class functions on the symmetric groups and their asymptotics. *Ann. Inst. Fourier (Grenoble)*, 63(4):1227–1262, 2013.
- [36] Persi Diaconis and Mehrdad Shahshahani. On the eigenvalues of random matrices. *J. Appl. Probab.*, 31A:49–62, 1994. Studies in applied probability.
- [37] Jean-Jil Duchamps, Jim Pitman, and Wenpin Tang. Renewal sequences and record chains related to multiple zeta sums. *arXiv preprint arXiv:1707.07776*, 2017.
- [38] P. Erdős and P. Turán. On some problems of a statistical group-theory. III. *Acta Math. Acad. Sci. Hungar.*, 18:309–320, 1967.
- [39] Steven N. Evans. Eigenvalues of random wreath products. *Electron. J. Probab.*, 7:no. 9, 15, 2002.
- [40] Warren J. Ewens. The sampling theory of selectively neutral alleles. *Theoret. Population Biology*, 3:87–112; erratum, *ibid.* 3 (1972), 240; erratum, *ibid.* 3 (1972), 376, 1972.
- [41] Valentin Féray. Asymptotic behavior of some statistics in Ewens random permutations. *Electron. J. Probab.*, 18:no. 76, 32, 2013.
- [42] Ronald A Fisher. The design of experiments. *Oliver and Boyd, Edinburgh*, 1935.
- [43] Alexander Gnedin, Alexander Iksanov, and Alexander Marynych. A generalization of the Erdős-Turán law for the order of random permutation. *Combin. Probab. Comput.*, 21(5):715–733, 2012.

- [44] Solomon W Golomb. Random permutations. *Bull. Amer. Math. Soc.*, 70:747, 1964.
- [45] Solomon W. Golomb. *Shift register sequences*. With portions co-authored by Lloyd R. Welch, Richard M. Goldstein, and Alfred W. Hales. Holden-Day, Inc., San Francisco, Calif.-Cambridge-Amsterdam, 1967.
- [46] V. Gontcharoff. Du domaine de l’analyse combinatoire. *Bull. Acad. Sci. URSS Sér. Math. [Izvestia Akad. Nauk SSSR]*, 8:3–48, 1944.
- [47] W. Gontcharoff. Sur la distribution des cycles dans les permutations. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 35:267–269, 1942.
- [48] B. M. Hambly, P. Keevash, N. O’Connell, and D. Stark. The characteristic polynomial of a random permutation matrix. *Stochastic Process. Appl.*, 90(2):335–346, 2000.
- [49] Fred M. Hoppe. Size-biased filtering of Poisson-Dirichlet samples with an application to partition structures in genetics. *J. Appl. Probab.*, 23(4):1008–1012, 1986.
- [50] Fred M. Hoppe. The sampling theory of neutral alleles and an urn model in population genetics. *J. Math. Biol.*, 25(2):123–159, 1987.
- [51] C. P. Hughes, J. P. Keating, and Neil O’Connell. On the characteristic polynomial of a random unitary matrix. *Comm. Math. Phys.*, 220(2):429–451, 2001.
- [52] Christopher Hughes, Joseph Najnudel, Ashkan Nikeghbali, and Dirk Zeindler. Random permutation matrices under the generalized Ewens measure. *Ann. Appl. Probab.*, 23(3):987–1024, 2013.
- [53] Axel Hultman. Permutation statistics of products of random permutations. *Adv. in Appl. Math.*, 54:1–10, 2014.
- [54] Svante Janson. Patterns in random permutations avoiding the pattern 132. *Combin. Probab. Comput.*, 26(1):24–51, 2017.
- [55] Tobias Johnson and Soumik Pal. Cycles and eigenvalues of sequentially growing random regular graphs. *Ann. Probab.*, 42(4):1396–1437, 2014.
- [56] Jean-Pierre Kahane. Sur le chaos multiplicatif. *Ann. Sci. Math. Québec*, 9(2):105–150, 1985.
- [57] Jon P. Keating and Nina C. Snaith. Random matrix theory and  $\zeta(1/2 + it)$ . *Comm. Math. Phys.*, 214(1):57–89, 2000.
- [58] Sergei Kerov, Grigori Olshanski, and Anatoly Vershik. Harmonic analysis on the infinite symmetric group. *Invent. Math.*, 158(3):551–642, 2004.
- [59] Sergei Kerov and Natalia Tsilevich. Stick breaking process generated by virtual permutations with ewens distribution. *Journal of Mathematical Sciences*, 87(6):4082–4093, 1997.

- [60] Serguei Kerov, Grigori Olshanski, and Anatoli Vershik. Harmonic analysis on the infinite symmetric group. A deformation of the regular representation. *C. R. Acad. Sci. Paris Sér. I Math.*, 316(8):773–778, 1993.
- [61] John Frank Charles Kingman. *Poisson processes*, volume 3 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1993. Oxford Science Publications.
- [62] Gaultier Lambert. Mesoscopic fluctuations for unitary invariant ensembles. *arXiv preprint arXiv:1510.03641*, 2015.
- [63] Gaultier Lambert, Dmitry Ostrovsky, and Nick Simm. Subcritical multiplicative chaos for regularized counting statistics from random matrix theory. *arXiv preprint arXiv:1612.02367*, 2016.
- [64] Mickaël Maazoun. On the brownian separable permuton. *arXiv preprint arXiv:1711.08986*, 2017.
- [65] Camille Male. Traffic distributions and independence: permutation invariant random matrices and the three notions of independence. *arXiv preprint arXiv:1111.4662*, 2011.
- [66] Madan Lal Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.
- [67] H. L. Montgomery. The pair correlation of zeros of the zeta function. pages 181–193, 1973.
- [68] Joseph Najnudel and Ashkan Nikeghbali. The distribution of eigenvalues of randomized permutation matrices. *Ann. Inst. Fourier (Grenoble)*, 63(3):773–838, 2013.
- [69] Joseph Najnudel and Ashkan Nikeghbali. On a flow of operators associated to virtual permutations. In *Séminaire de Probabilités XLVI*, volume 2123 of *Lecture Notes in Math.*, pages 481–512. Springer, Cham, 2014.
- [70] Mihail G. Neagu. Asymptotic freeness of random permutation matrices from Gaussian matrices. *J. Ramanujan Math. Soc.*, 20(3):189–213, 2005.
- [71] Mihail G. Neagu. Asymptotic freeness of random permutation matrices with restricted cycle lengths. *Indiana Univ. Math. J.*, 56(4):2017–2049, 2007.
- [72] Alexandru Nica. On the number of cycles of given length of a free word in several random permutations. *Random Structures Algorithms*, 5(5):703–730, 1994.
- [73] Ashkan Nikeghbali and Dirk Zeindler. The generalized weighted probability measure on the symmetric group and the asymptotic behavior of the cycles. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(4):961–981, 2013.
- [74] Grigori Olshanski. Random permutations and related topics. In *The Oxford handbook of random matrix theory*, pages 510–533. Oxford Univ. Press, Oxford, 2011.

- [75] G. P. Patil and C. Taillie. Diversity as a concept and its implications for random communities. *Bull. Inst. Internat. Statist.*, 47(2):497–515, 551–558, 1977. With discussion.
- [76] Jim Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard.
- [77] Jim Pitman and Wenpin Tang. Regenerative random permutations of integers. *arXiv preprint arXiv:1704.01166*, 2017.
- [78] Jim Pitman and Yuri Yakubovich. Extremes and gaps in sampling from a GEM random discrete distribution. *Electron. J. Probab.*, 22:Paper No. 44, 26, 2017.
- [79] E. M. Rains. Increasing subsequences and the classical groups. *Electron. J. Combin.*, 5:Research Paper 12, 9, 1998.
- [80] C. Schensted. Longest increasing and decreasing subsequences. *Canad. J. Math.*, 13:179–191, 1961.
- [81] L. A. Shepp and S. P. Lloyd. Ordered cycle lengths in a random permutation. *Trans. Amer. Math. Soc.*, 121:340–357, 1966.
- [82] Julia Storm and Dirk Zeindler. The order of large random permutations with cycle weights. *Electron. J. Probab.*, 20:no. 126, 34, 2015.
- [83] Craig A. Tracy and Harold Widom. Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.*, 159(1):151–174, 1994.
- [84] Craig A. Tracy and Harold Widom. Random unitary matrices, permutations and Painlevé. *Comm. Math. Phys.*, 207(3):665–685, 1999.
- [85] Natalia Tsilevich. Distribution of cycle lengths of infinite permutations. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 223(Teor. Predstav. Din. Sistemy, Kombin. i Algoritm. Metody. I):148–161, 339, 1995.
- [86] G. A. Watterson. The sampling theory of selectively neutral alleles. *Advances in Appl. Probability*, 6:463–488, 1974.
- [87] Kelly Wieand. Eigenvalue distributions of random permutation matrices. *Ann. Probab.*, 28(4):1563–1587, 2000.
- [88] Kelly Wieand. Eigenvalue distributions of random unitary matrices. *Probab. Theory Related Fields*, 123(2):202–224, 2002.
- [89] Kelly Wieand. Permutation matrices, wreath products, and the distribution of eigenvalues. *J. Theoret. Probab.*, 16(3):599–623, 2003.
- [90] John Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, pages 32–52, 1928.

- [91] A. L. Yakymiv. Random  $A$ -permutations: convergence to a Poisson process. *Mat. Zametki*, 81(6):939–947, 2007.
- [92] Dirk Zeindler. Permutation matrices and the moments of their characteristic polynomial. *Electron. J. Probab.*, 15:no. 34, 1092–1118, 2010.
- [93] Dirk Zeindler. Central limit theorem for multiplicative class functions on the symmetric group. *J. Theoret. Probab.*, 26(4):968–996, 2013.
- [94] James Y. Zhao. Universality of asymptotically Ewens measures on partitions. *Electron. Commun. Probab.*, 17:no. 16, 11, 2012.
- [95] A. Zygmund. *Trigonometric series. Vol. I, II.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, third edition, 2002. With a foreword by Robert A. Fefferman.

# Spectre de matrices de permutation aléatoires

## RÉSUMÉ

Dans cette thèse, nous nous intéressons à des matrices aléatoires en lien avec des permutations. Nous abordons l'étude de leurs spectres de plusieurs manières, et à différentes échelles d'observation.

Dans un premier temps, nous prolongeons l'étude de Wieand à propos des nombres de valeurs propres appartenant à certains arcs fixés du cercle unité. Pour cela nous tirons parti des travaux réalisés par Ben Arous et Dang sur les statistiques linéaires du spectre de matrices de permutation pour une famille de lois à un paramètre englobant le cas de la loi uniforme sur le groupe symétrique, appelée famille des lois d'Ewens. Une partie innovante de notre travail réside dans la généralisation à des arcs non nécessairement fixés. Nous obtenons en effet des résultats similaires en autorisant les longueurs des arcs à décroître lentement vers zéro avec la taille des matrices.

Dans un deuxième temps, nous regardons le spectre à échelle microscopique. En nous inspirant des travaux de Najnudel et Nikeghbali en rapport avec la convergence de mesures empiriques des angles propres normalisés, nous commençons par donner un sens à la convergence en terme de comptages de points sur des intervalles fixés. À partir du processus ponctuel limite, nous montrons que le nombre de points dans un intervalle a des fluctuations asymptotiquement gaussiennes lorsque la longueur de l'intervalle tend vers l'infini.

Enfin, nous adaptons certains résultats de Chhaibi, Najnudel et Nikeghbali sur le polynôme caractéristique de matrices du CUE à échelle microscopique, et les développons dans notre cadre. De manière analogue mais avec d'autres techniques de preuves, nous obtenons des convergences des polynômes caractéristiques vers des fonctions entières, et cela pour une grande famille de lois pour le tirage des permutations, incluant les lois d'Ewens.

## ABSTRACT

In this thesis, our goal is to study random matrices related to permutations. We tackle the study of their spectra in various ways, and at different scales.

First, we extend the work of Wieand about the numbers of eigenvalues lying in some fixed arcs of the unit circle. We take advantage of the results of Ben Arous and Dang on the linear statistics of the spectrum of permutation matrices for a one-parameter family of deformations of the uniform law on the symmetric group, called Ewens' measures. One of the most innovative parts of our work is the generalization to non-fixed arcs. Indeed we get similar results when we let the lengths of the arcs decrease to zero slower than  $1/n$ .

Then, we look at the spectrum at microscopic scale. Inspired by the work of Najnudel and Nikeghbali about the convergence of empirical measures of rescaled eigenangles, we give a meaning to the convergence in terms of indicator functions of intervals. From the limiting point process, we show that the number of points in any interval is asymptotically normal as the length of the interval goes to infinity.

Finally, we adapt some results of Chhaibi, Najnudel and Nikeghbali on the characteristic polynomial of the CUE at microscopic scale, and develop them in our framework. Analogously but with different techniques of proof, we get that the characteristic polynomials converge to entire functions, and this for a large family of laws including the Ewens' measures.